Cluster Analysis of Air Quality Data for CCOS Study Domain

Scott Beaver

Ahmet Palazoglu, P.I.

University of California, Davis

Dept. Chemical Engineering & Materials Science

CCOS Technical Committee Meeting

Cal/EPA Building, 1001 I St., Sacramento, CA

Friday 15 December 2006

Overview

- Intro to cluster analysis
  - Clustering for air quality analysis
  - CCOS project overview
- Completed Bay Area analysis
  - Description of monitoring network
  - O_3 spatial field clustering
  - Wind field clustering
  - Wind field cluster sequencing
- Initial San Joaquin Valley analysis
  - Selection of monitoring network
  - North SJV wind field clustering
- Future work and recommendations

Cluster Analysis

<u>Cluster Analysis goal</u>:

Given *n* observations on *m* variables…

… partition the *n* observations into *k* clusters:

- Observations in same cluster are "similar"
- Clusters themselves are sufficiently "different"

- Clustering for air quality data
  - Clusters are groups of days
    - Days in same cluster share chemical and/or meteorological features
    - Each cluster captures unique set of features
  - Discriminate various features by comparing between clusters
- Unsupervised statistical method
  - No *a priori* knowledge of set of states required
  - Must formulate problem such that patterns relevant to ozone

Cluster Analysis for CCOS

- Study Domain
  - 6 CCOS air basins
    - San Francisco Bay Area
    - SJV: split into North, Central, & South
    - Sacramento Valley
    - Mountain Counties
  - 1996-2005 ozone seasons (1 May – 31 October)
- Intra-basin analysis
  - Recurring ozone spatial patterns
    - Mesoscale meteorological features and emissions scenarios
  - Recurring diurnal wind field patterns
    - Synoptic influences and mesoscale dispersion patterns
- Inter-basin analysis
  - Determine how similar features affect multiple basins

Ozone Field Clustering

- Input data:
  - Include only days with max. 8-hr O_3 > threshold
    - Thresholds of 85 & 70 ppb
  - Daily maximum 8-hr $[O_3]$ at *m* monitoring locations
- Clusters:
  - Days having similar spatial distribution for O_3
- Patterns:
  - Mesoscale met. and/or emissions characteristics
    - Clusters are not necessarily a physical regime
    - Presence/absence of characteristics from certain clusters allows identification of mechanisms
- Goal:
  - Determine various mechanisms for ozone buildup

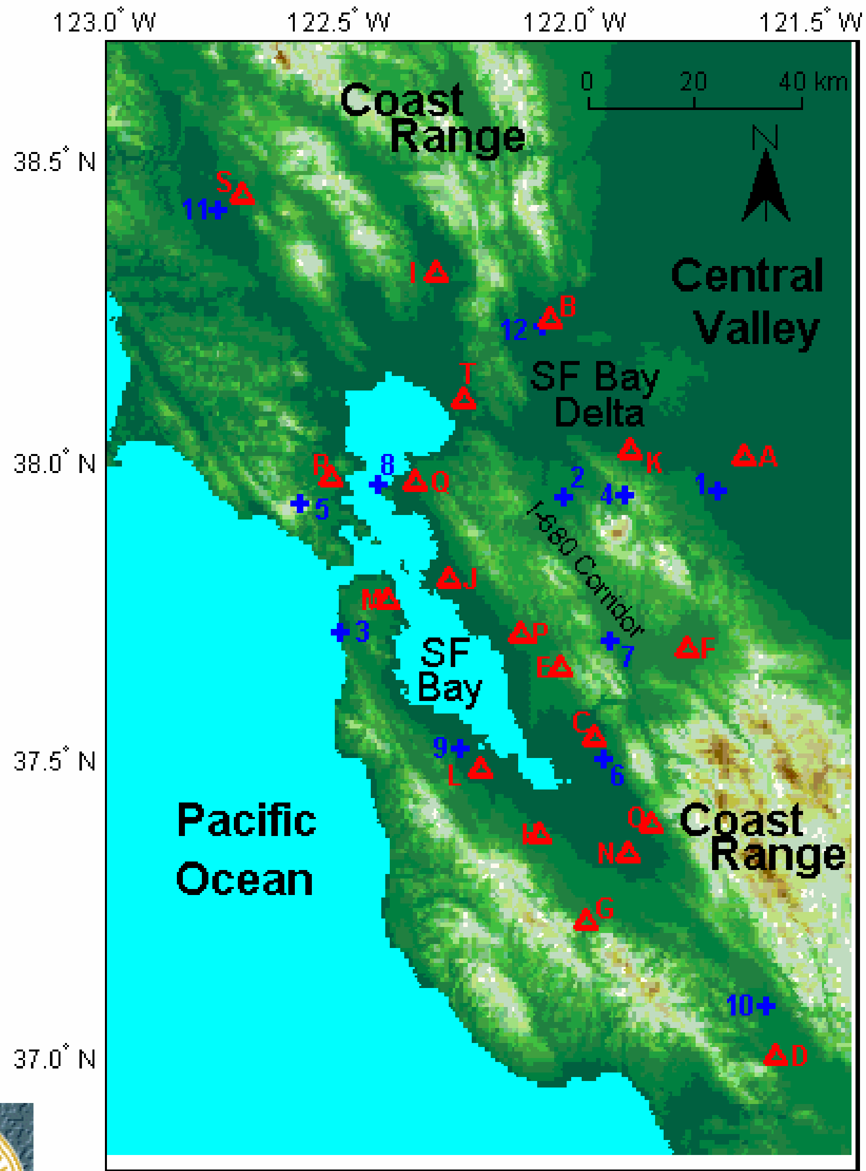
Wind Field Clustering

- Input Data:
  - All days in ozone season
  - 24 hourly *u* & *v* wind components at *m* monitoring locations
- Clusters:
  - Days with similar diurnal cycle for wind field
- Patterns:
  - Directly reveals mesoscale dispersion patterns
  - Indirectly reveals synoptic influences
    - Each cluster associated with a particular synoptic regime
- Goal:
  - Determine effect of synoptic meteorology on ozone

Uses for Cluster Analysis Results

- Meteorological Classification
  - Identify met scenarios for ozone buildup
  - Identify/confirm transport mechanisms
- Episode Selection for AQM simulation
  - Assess representativeness of simulated conditions
- Trend Analysis
  - Account for different emissions scenarios after normalizing for meteorology
- Design of future field programs
  - Determine limitations and redundancies for monitoring networks

San Francisco Bay Area



Meteorological Stations (+)

1	Bethel Island	-2 m
2	Concord	24 m
3	Fort Funston	57 m
4	Kregor Peak	577 m
5	Mt. Tamalpais	762 m
6	NUMMI	9 m
7	Pleasanton	99 m
8	Pt. San Pablo	70 m
9	San Carlos	1 m
10	San Martin	85 m
11	Santa Rosa	29 m
12	Suisun	5 m

Air Quality Stations (Δ or +)

A	Bethel Island*	6 m
2	Concord*	26 m
B	Fairfield	4 m
C	Fremont*	24 m
D	Gilroy	55 m
E	Hayward	288 m
F	Livermore*	145 m
G	Los Gatos	186 m
H	Mountain View	24 m
I	Napa*	22 m
J	Oakland	7 m
K	Pittsburg*	9 m
L	Redwood City*	9 m
M	San Francisco*	5 m
N	San Jose*	24 m
O	San Jose (East)	63 m
P	San Leandro	36 m
10	San Martin	87 m
Q	San Pablo*	15 m
R	San Rafael*	11 m
S	Santa Rosa*	49 m
T	Vallejo*	30 m

* Station monitors NO_x

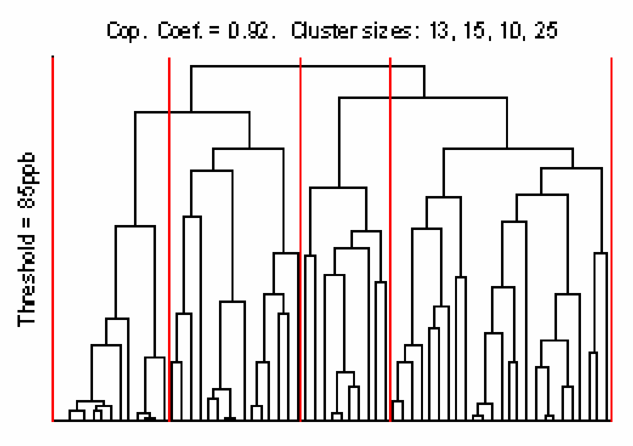
SFBA Ozone Clustering

- Daily max. 8-hr [O_3] from 22 locations
- Ozone clustering for 2 data sets:

1. Threshold = 85 ppb for 8-hr [O_3]
    - 63 days from 1996—2004
2. Threshold = 70 ppb for 8-hr [O_3]
    - 199 days from 1996—2004
        - 63 "old" days with $[O_3]_{max} > 85$
        - 136 "new" days with $70 < [O_3]_{max} < 85$
    - Do new & old days share same features?

*Beaver and Palazoglu, 2006: A cluster aggregation scheme for ozone episode selection in the San Francisco, CA Bay Area. *Atmospheric Environment*, **40**, 713—725.

85 ppb Threshold Results

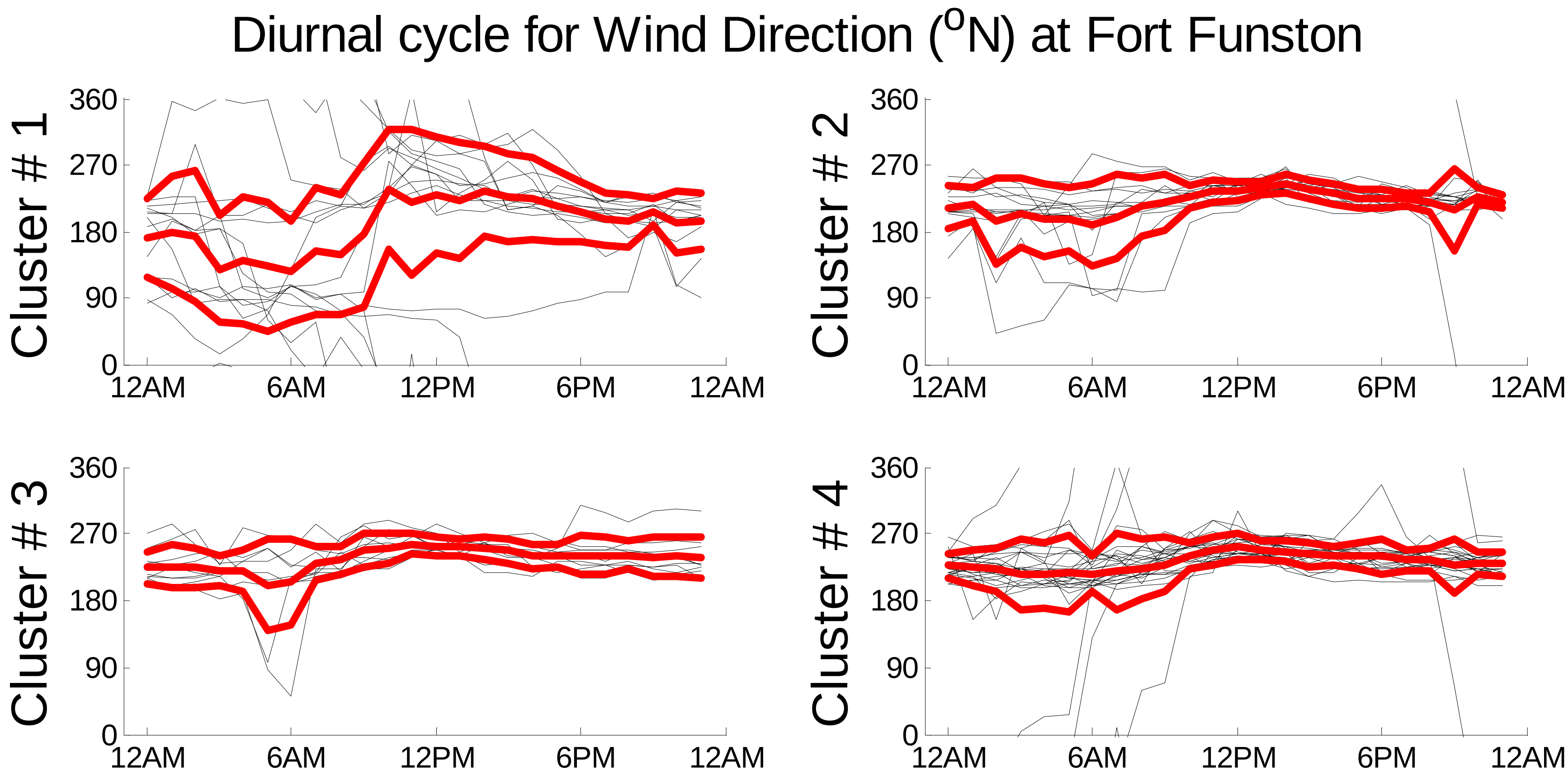


Above: Dendrogram showing 4-cluster solution.

Right: Fraction of days exceeding 85 ppb per subregion/location in each cluster.

	#1	#2	#3	#4
	13 Days	15 Days	10 Days	25 Days
South Bay	**0.46**	**0.27**	**0.10**	**0.20**
Fremont	0.08	0.00	0.00	0.04
Hayward	0.15	0.00	0.00	0.04
Los Gatos	0.23	0.27	0.10	0.20
Mountain View	0.15	0.07	0.00	0.00
Redwood City	0.00	0.00	0.00	0.00
San Leandro	0.00	0.00	0.00	0.00
E. San Jose	0.00	0.00	0.00	0.00
San Jose	0.00	0.00	0.00	0.04
East Bay	**0.00**	**0.47**	**0.40**	**0.68**
Bethel	0.00	0.13	0.20	0.60
Fairfield	0.00	0.20	0.00	0.28
Pittsburg	0.00	0.07	0.10	0.20
Concord	0.00	0.47	0.20	0.48
Livermore Valley	**0.08**	**0.80**	**0.70**	**0.72**
Livermore	0.08	0.80	0.70	0.72
Santa Clara Valley	**0.77**	**0.20**	**0.80**	**0.32**
Gilroy	0.46	0.00	0.50	0.12
San Martin	0.62	0.20	0.70	0.32

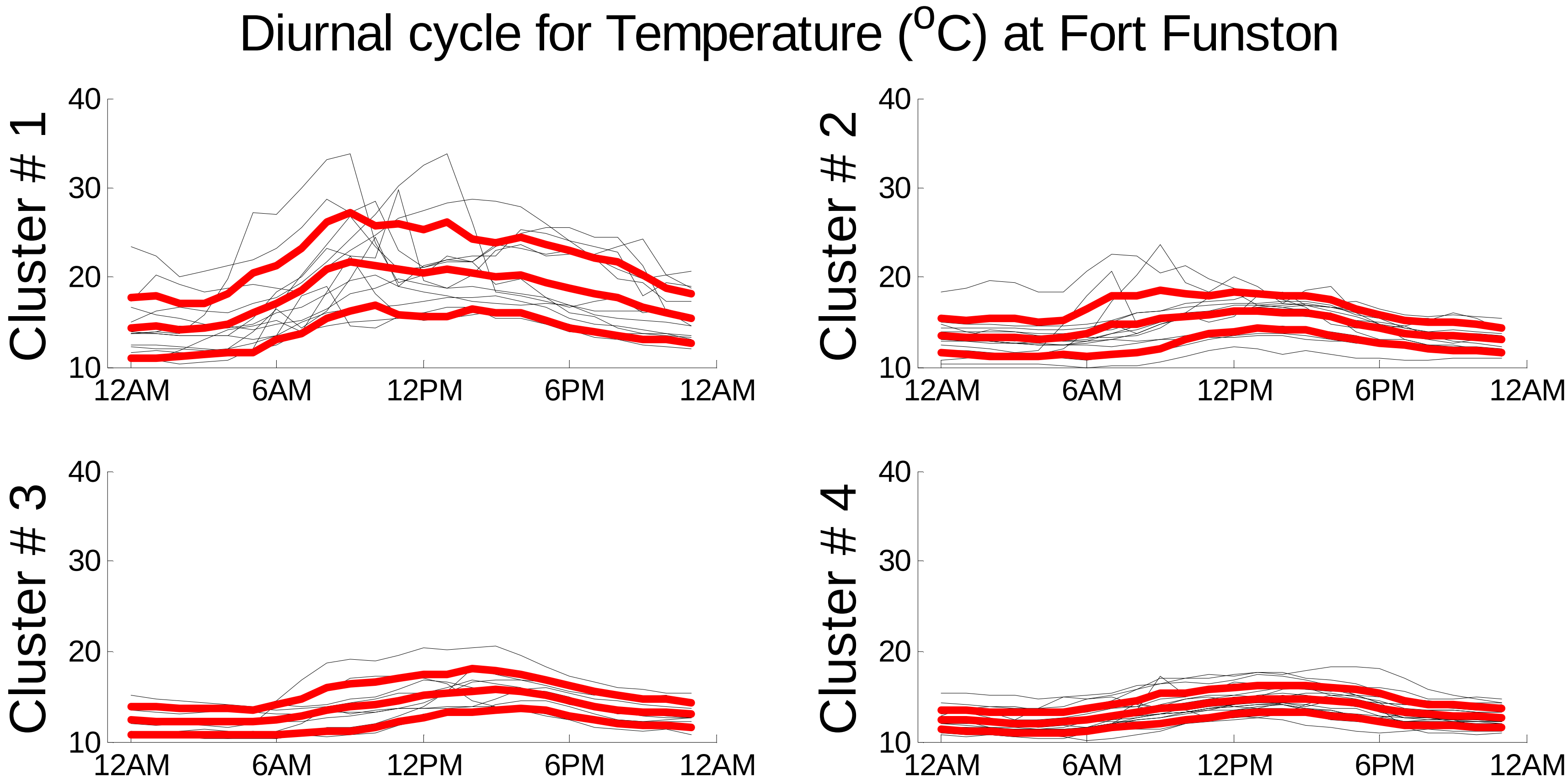
South Bay -- AM Flow Reversal



#1 exhibits reversal of flow from westerly to easterly origin.

Elevated South Bay ozone levels for #1, with variable peak location.

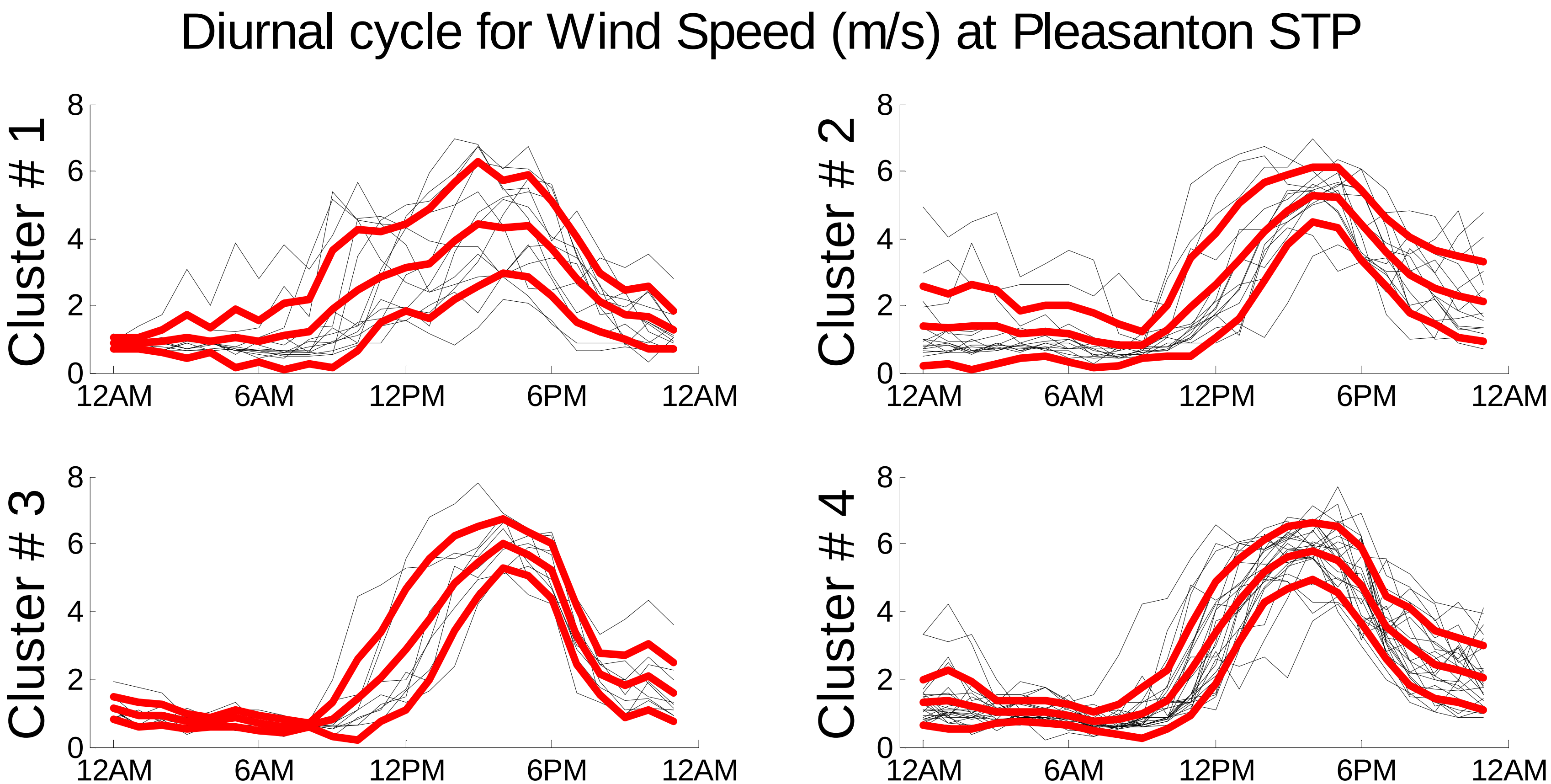
Flow Reversal & Coastal Temp



Elevated coastal temperatures for #1.

Indicates lack of marine flow into Bay Area.

Livermore Valley -- Seabreeze

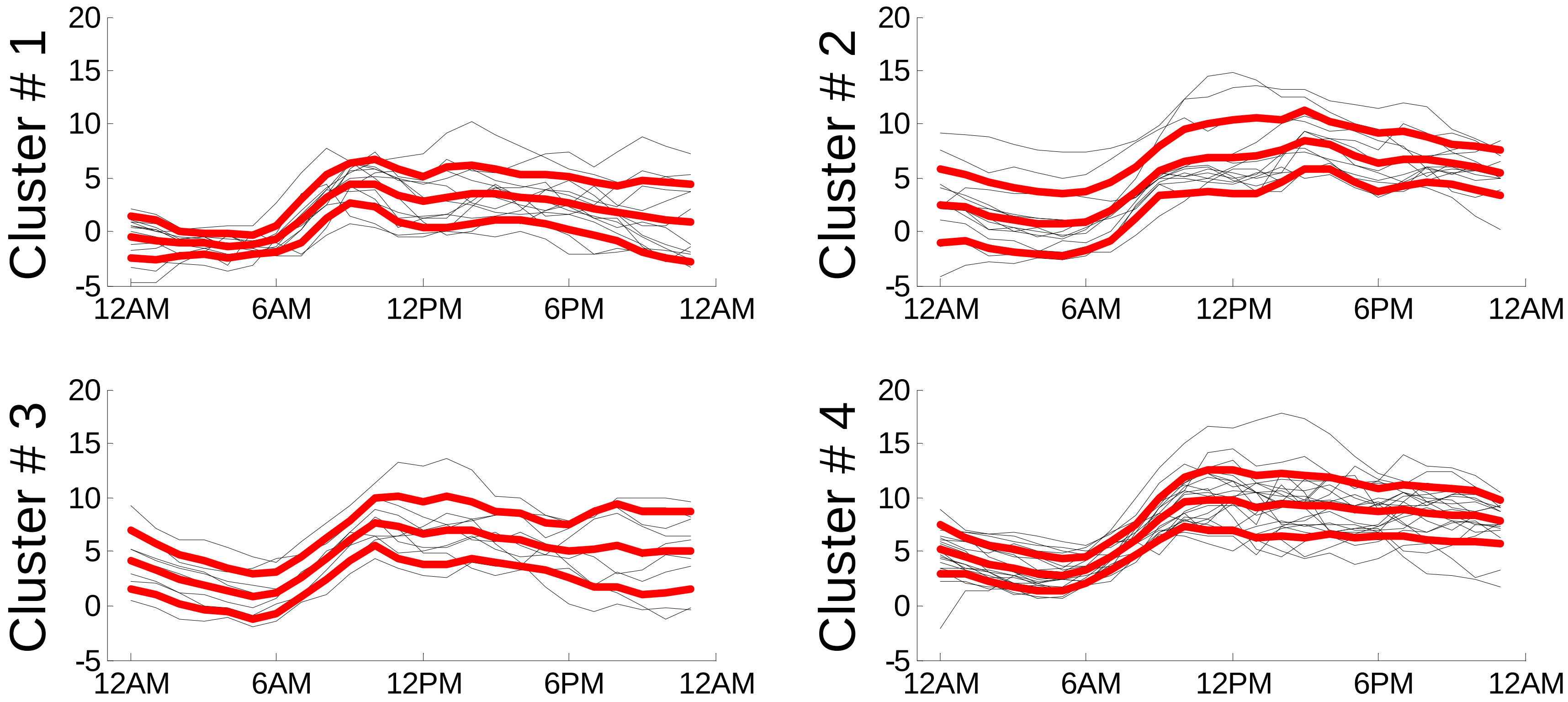


#1 has low wind speed in LV, indicating lack of seabreeze.

Other 3 clusters have seabreeze and high ozone levels.

LV Seabreeze & Temp Gradient

Diurnal cycle for Pleasanton STP - Pt. San Pablo Temp difference (^oC) at



#1 has small temperature gradient between LV and mouth of Bay.

Seabreeze is suppressed for #1; no transport into LV.

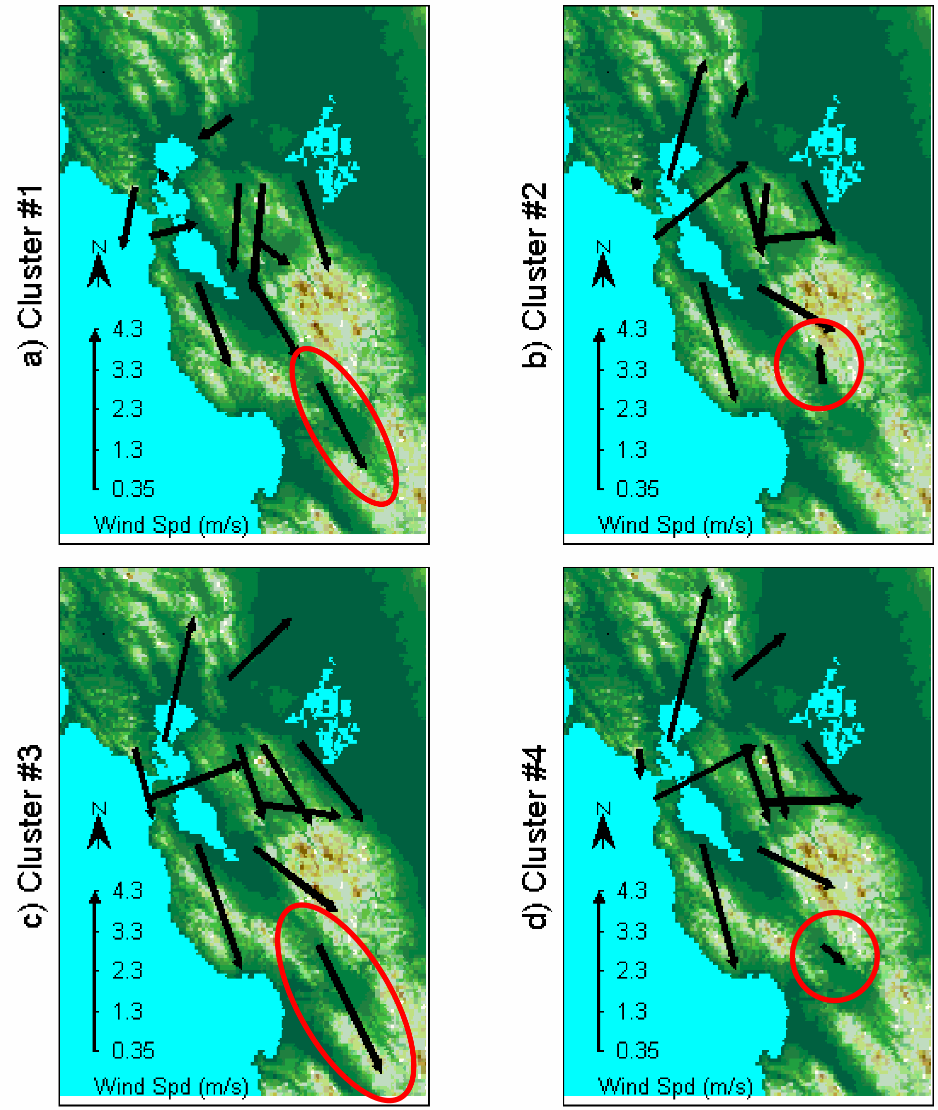
Department of Chemical Engineering and Materials Science

Santa Clara Valley -- Seabreeze

<u>Mean noontime wind fields:</u>

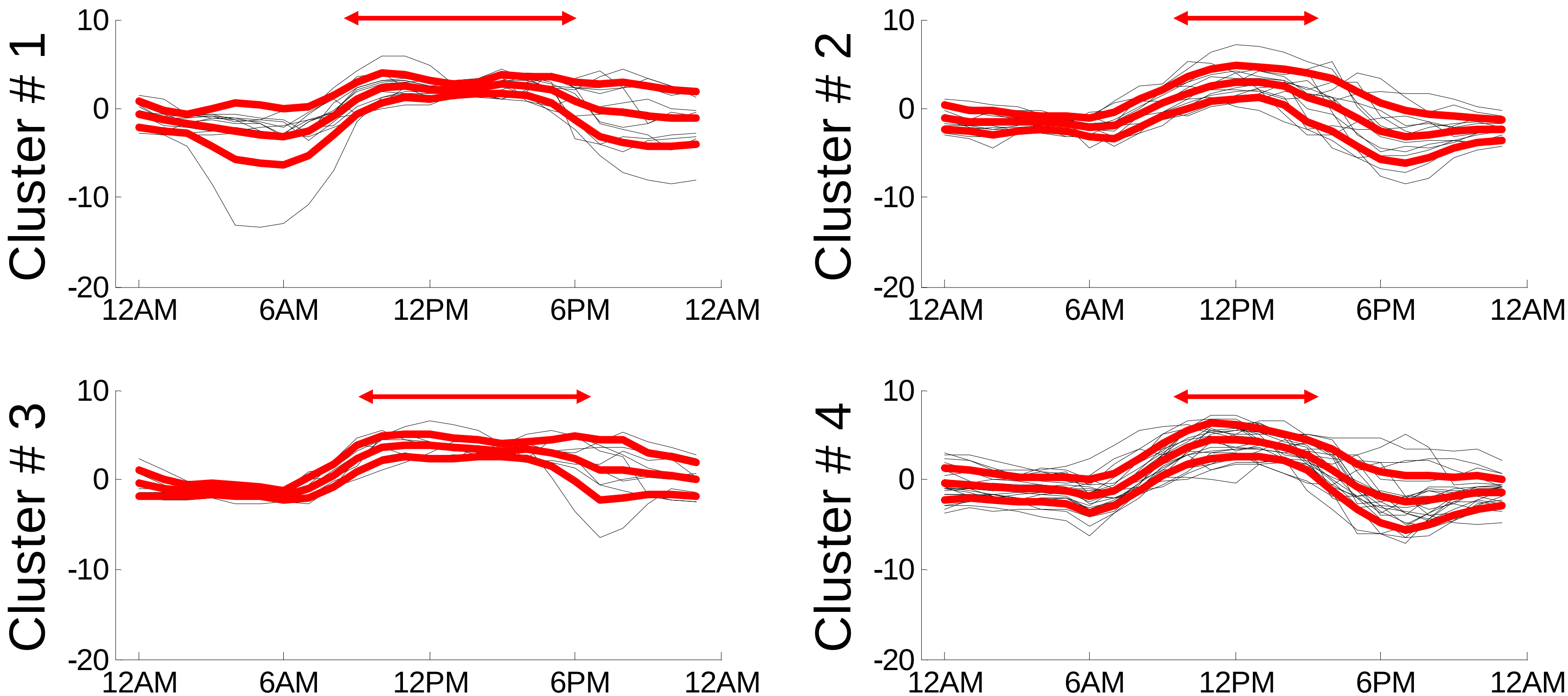
#1 and #3 have frequent SCV exceedances, while #2 & #4 have occasional exceedances.

#1 and #3 only exhibit transport into SCV from South Bay, as noted at San Martin (shown with red).



SCV Seabreeze & Temp Gradient

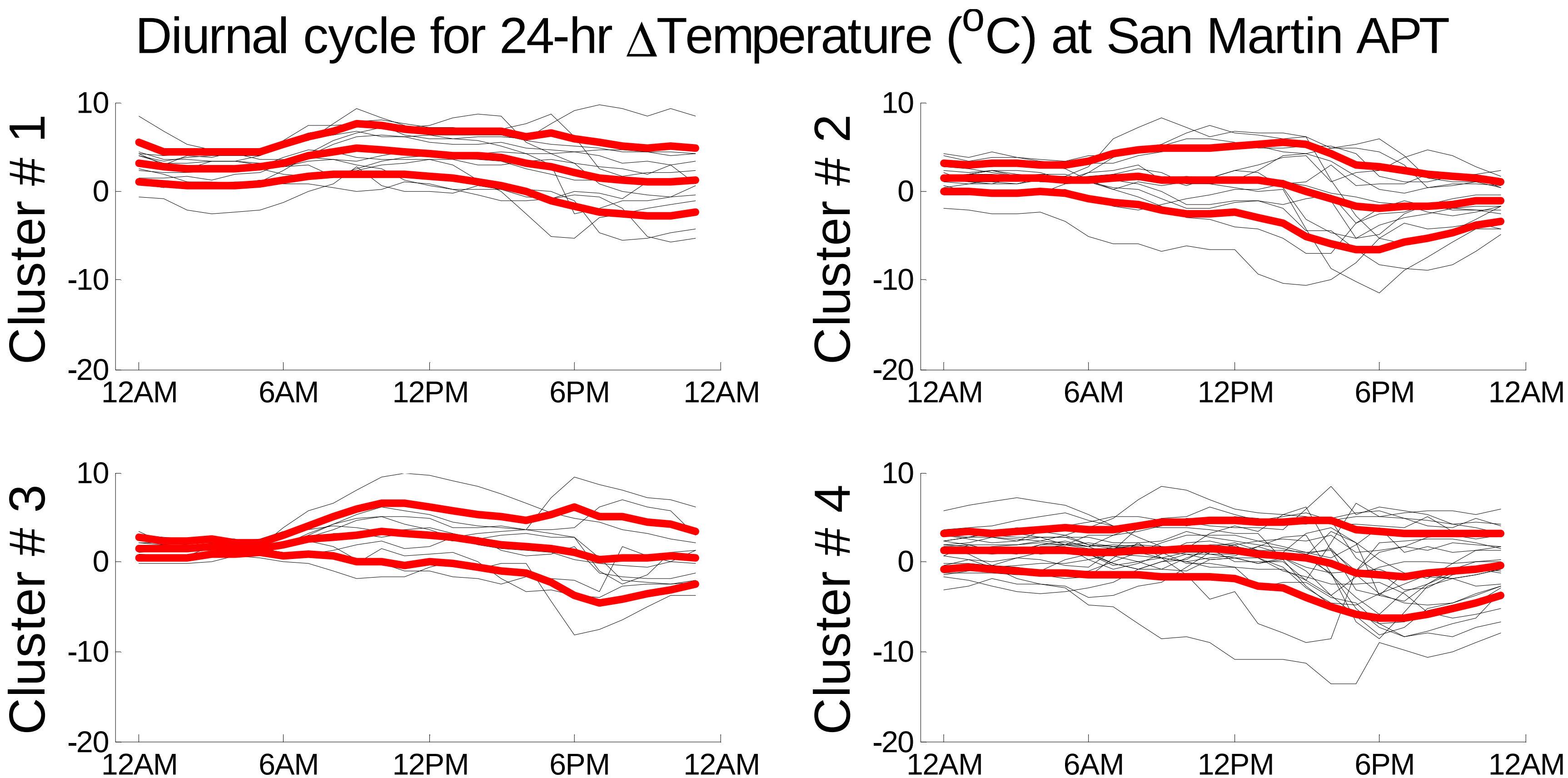
Diurnal cycle for San Martin APT - NUMMI Temp difference (^{o}C) at



#1 and #3 have seabreeze starting earlier in day and lasting longer into the evening relative clusters #2 and #4.

Magnitude of temperature gradient is less important.

Santa Clara Valley Temp Rise



#1 and #3 days are considerably warmer than previous day, especially into the early evening hours.

Changing conditions for #1 and #3 associated with episodes.

East Bay -- Weekday Effect

- Clusters #2 and #4 share similar met features and O_3 spatial distribution
  - No flow reversal
  - No SCV seabreeze; LV seabreeze present
- But #4 has unusually high East Bay O_3

Cluster:	#1	#2	#3	#4	Total
# Days:	13	15	10	25	63
% Weekend:	46%	73%	50%	16%	41%
% LV Exceedances	8%	80%	70%	72%	60%
% EB Exceedances	0%	47%	40%	68%	44%

Weekdays (Mon to Fri) favor East Bay exceedances.

Weekends (Sat and Sun) favor Livermore Valley exceedances.

70 ppb Threshold Results

- Now 199 days w/ $[O_3]_{max} > 70$
  - 63 “old” days with $[O_3]_{max} > 85$
  - 136 “new” days with $70 < [O_3]_{max} < 85$
- Similar 4-cluster solution is found
  - Flow reversal
    - Becomes more prevalent, largest category for new days
    - Biased toward ends of ozone season; rarely occurs in July or August.
  - East Bay weekend effect
    - Becomes more prevalent, 2nd largest category for new days
    - Larger disparity between East Bay and Livermore Valley O_3 levels
  - Livermore and Santa Clara Valleys
    - Similar seabreeze transport associated with elevated O_3

Summary

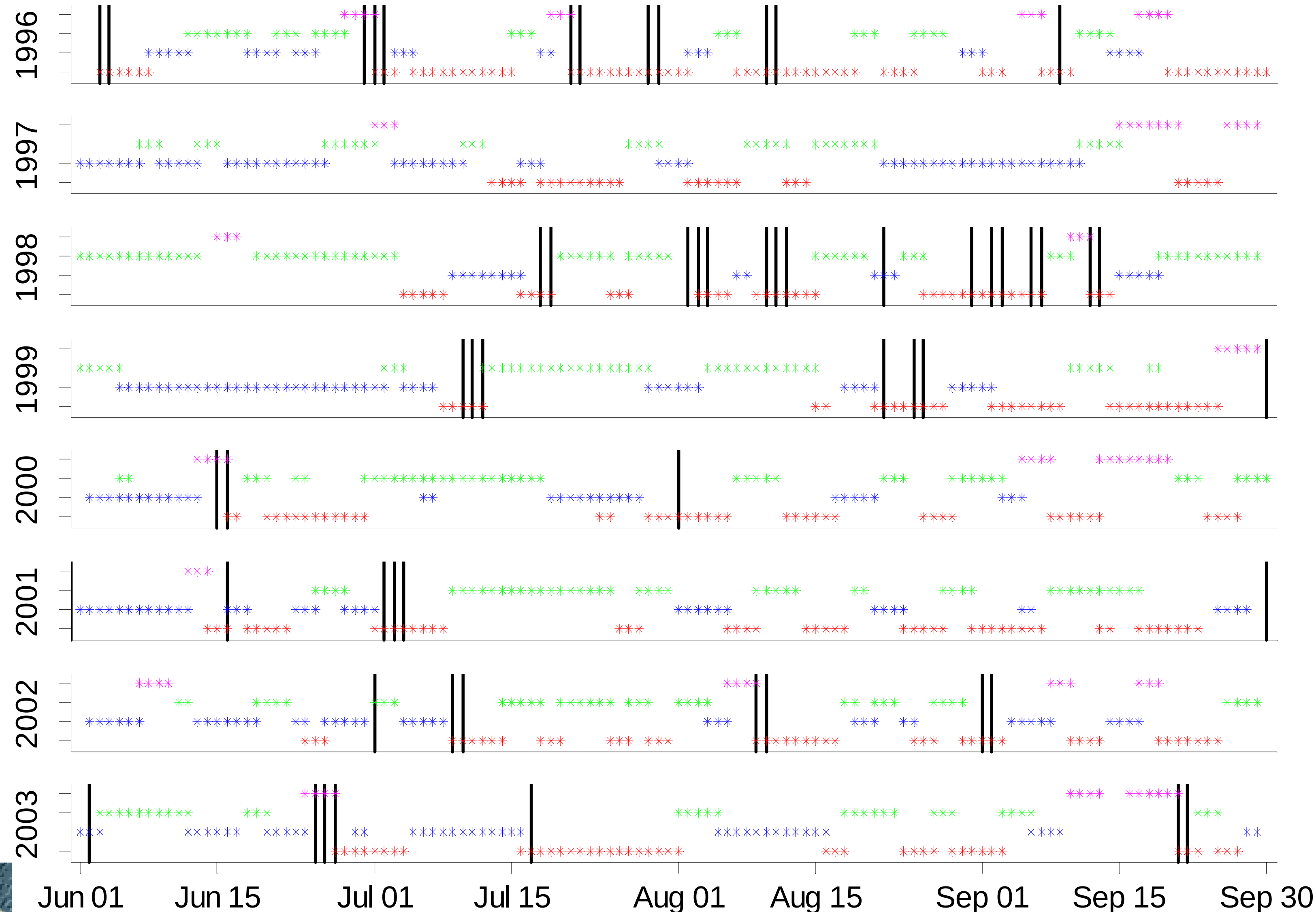
- South Bay
  - Seasonal Flow Reversal
- Livermore Valley
  - Seabreeze transport
  - Weekend effect
- Santa Clara Valley
  - Seabreeze transport
  - Abruptly changing conditions
- East Bay
  - Weekday affect
- Similar mechanisms observed for most days with $[O_3]_{max} > 70$ ppb

SFBA Wind Field Clustering

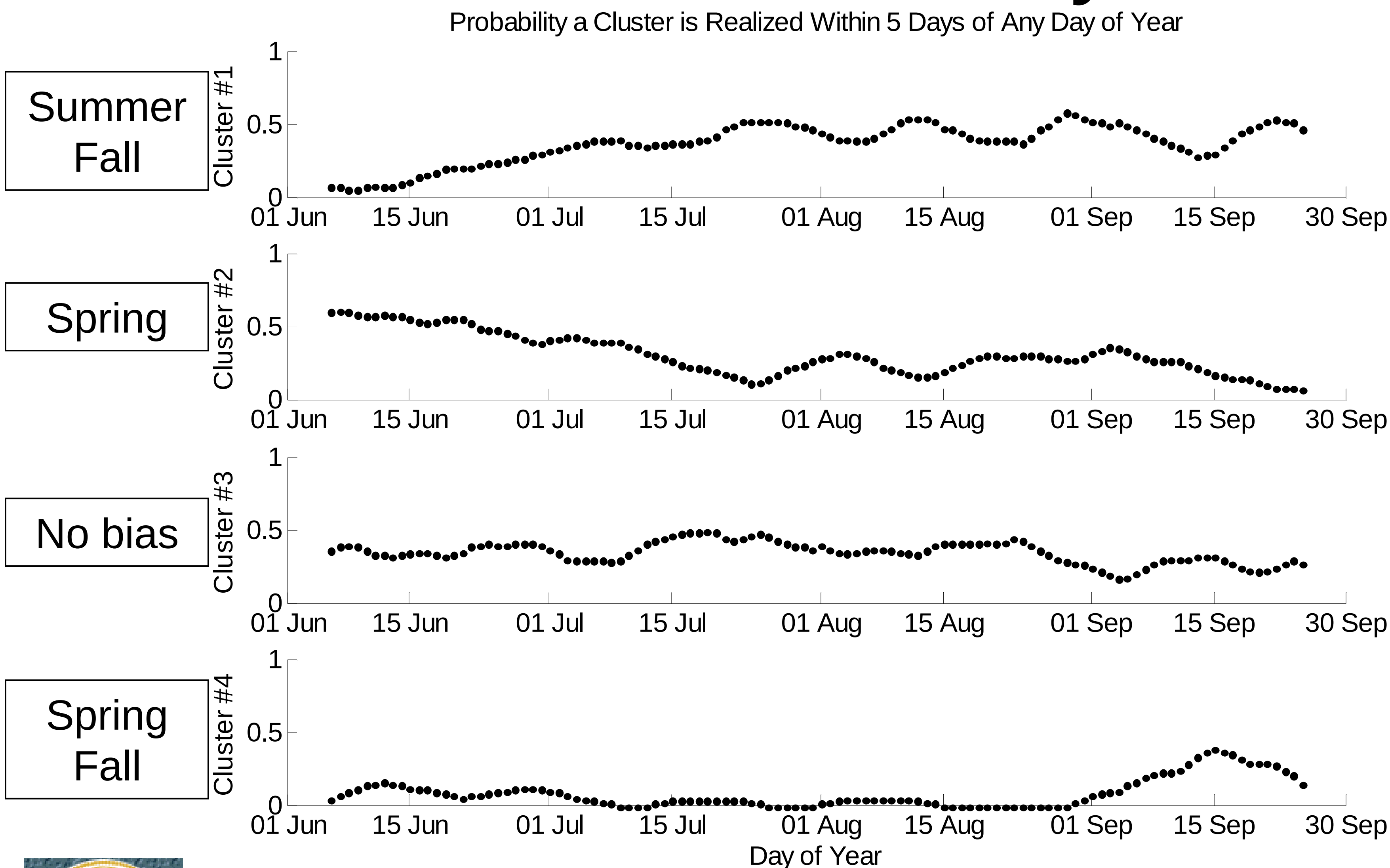
- Hourly *u* and *v* wind components from 11 locations
- Cluster every day of ozone season
  - 1996--2004
  - 1 June through 30 September
  - May and October excluded due to low frequency of Bay Area episodes
- Group days with similar 24-hr wind field evolution

*Beaver and Palazoglu, 2006: Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality. *Journal of Applied Meteorology and Climatology*, **45**, 1710—26.

Wind Field Clustering Results



Mild Seasonality



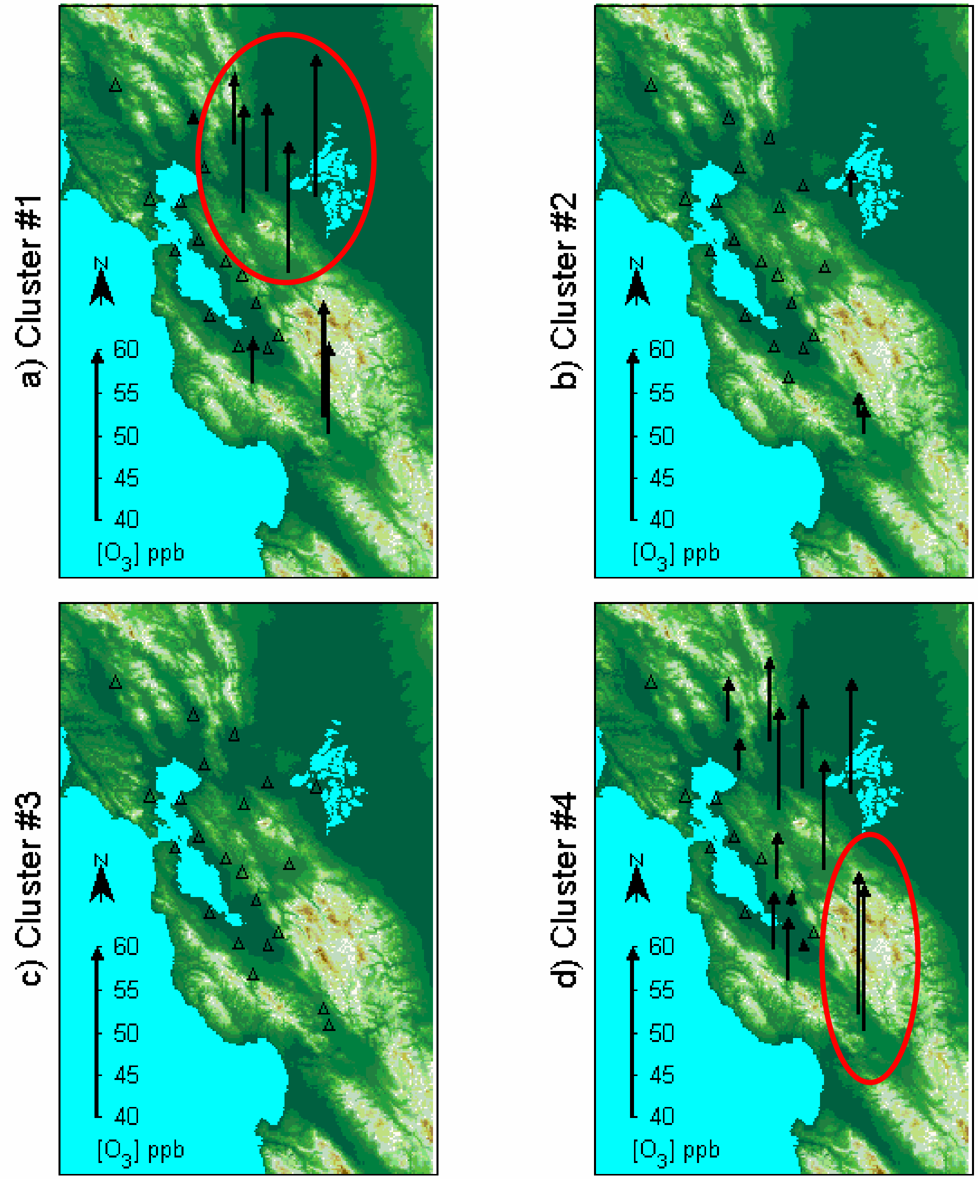
Ozone Response

#1: 353 days
Elevated ozone, highest in Livermore Valley & East Bay

#2: 309 days
Low ozone levels

#3: 341 days
Lowest ozone levels

#4: 86 days
Elevated ozone, highest in Santa Clara Valley



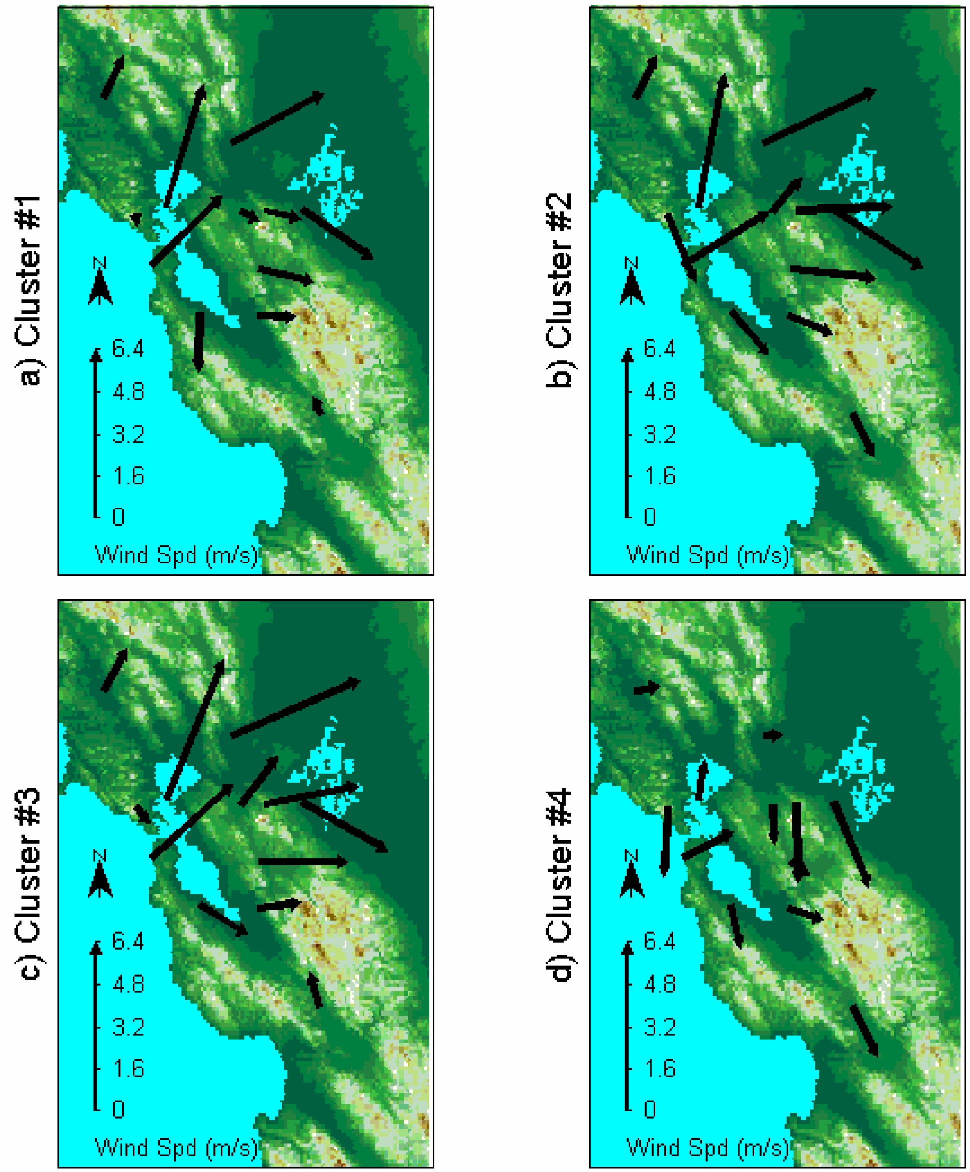
Midday Mesoscale Flow Patterns

#1: Weak marine layer penetration; flow along I-680 corridor into Livermore Valley

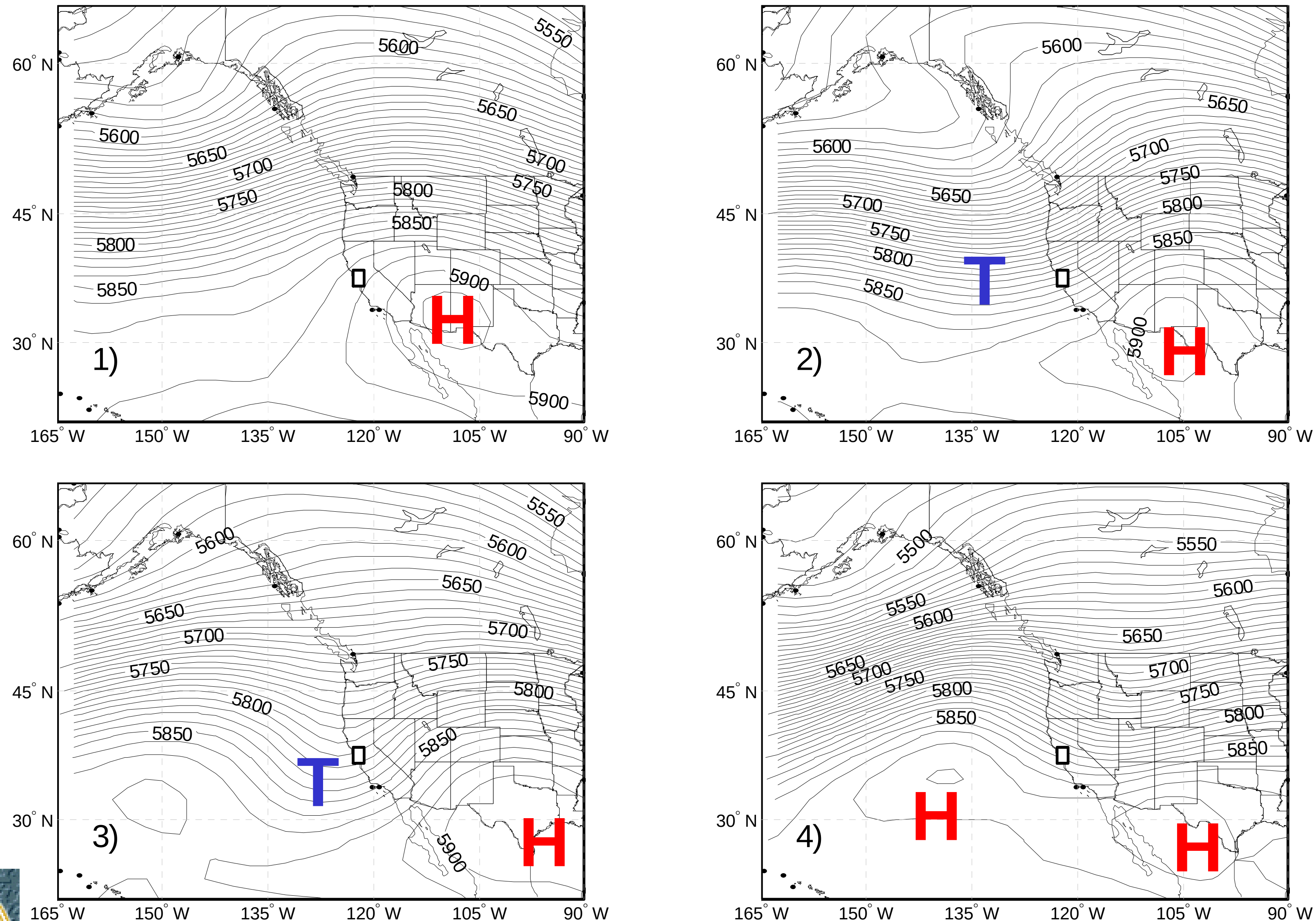
#2: Marine flow enters mouth of Bay and channels through Bay Area

#3: Stronger channeling marine flow than #2

#4: Little marine intrusion; northerly shift in wind direction

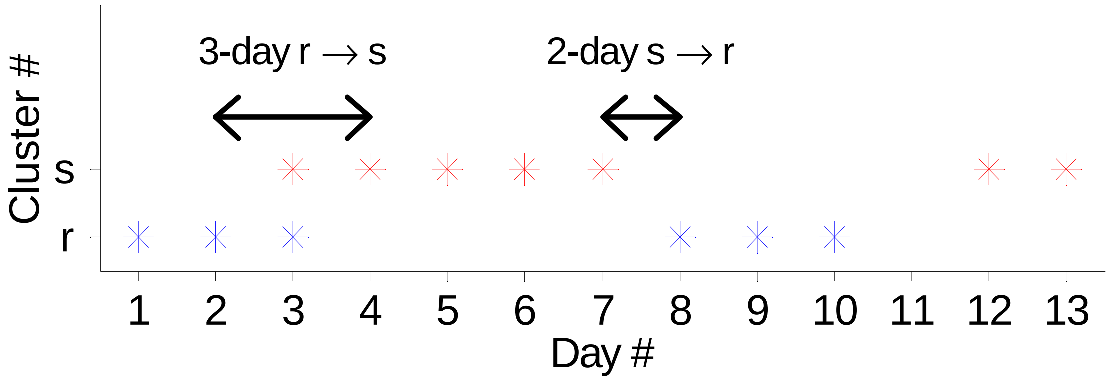


Synoptic Influences



SFBA Wind Field Sequencing

- Cluster transitions imply change in synoptic state
  - Binomial statistics: Determine if transition *r*→*s* is “favored”, “disfavored” or neither (random).
- Cluster Persistency
  - Run length for cluster realizations



*Beaver, Palazoglu and Tanrikulu, 2006: Cluster sequencing to analyze synoptic transitions affecting regional ozone. Submitted to *Journal of Applied Meteorology and Climatology*.

Favored/Disfavored Transitions

Null Hypothesis: State transitions are random.

	to #1	to #2	to #3	to #4
from #1	---	0.35	0.41	0.13
from #2	0.38	---	0.39	0.12
from #3	0.40	0.35	---	0.13
from #4	0.31	0.28	0.32	---

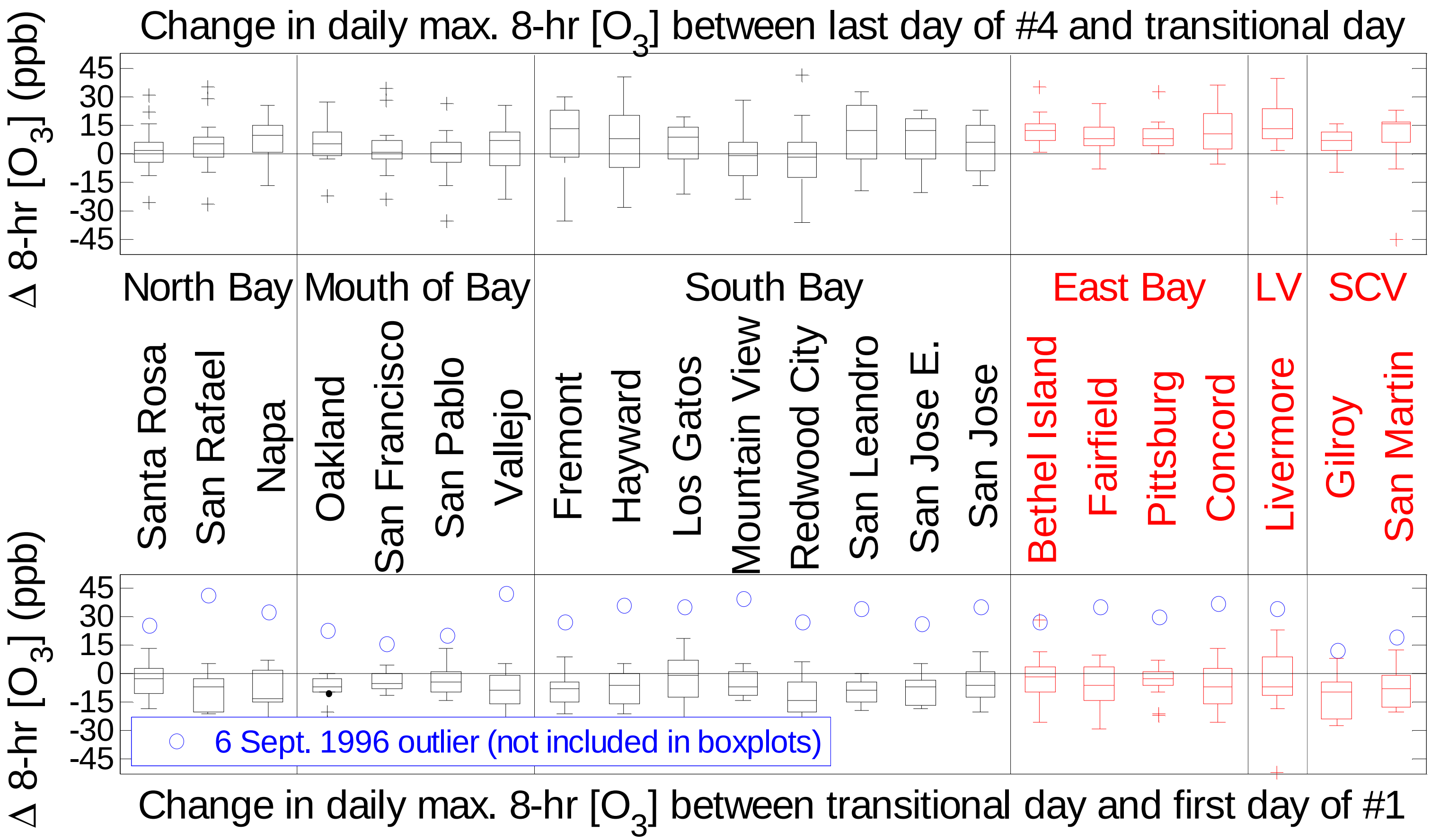
Test statistics with $\alpha = 0.05$ confidence bounds.

	to #1	to #2	to #3	to #4
from #1	---	0.30 ± 0.11	0.60 ± 0.12	0.05 ± 0.05
from #2	0.39 ± 0.12	---	0.39 ± 0.12	0.21 ± 0.10
from #3	0.40 ± 0.12	0.44 ± 0.12	---	0.09 ± 0.07
from #4	0.64 ± 0.19	0.12 ± 0.14	0.16 ± 0.15	---

Blue = Favored Transition

Red = Disfavored Transition

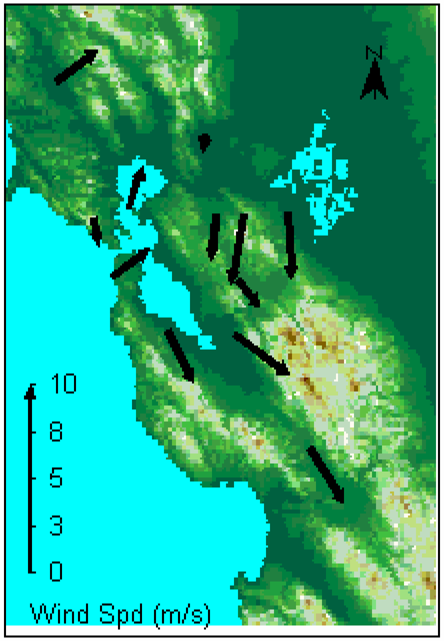
#4 → #1 Transition O_3



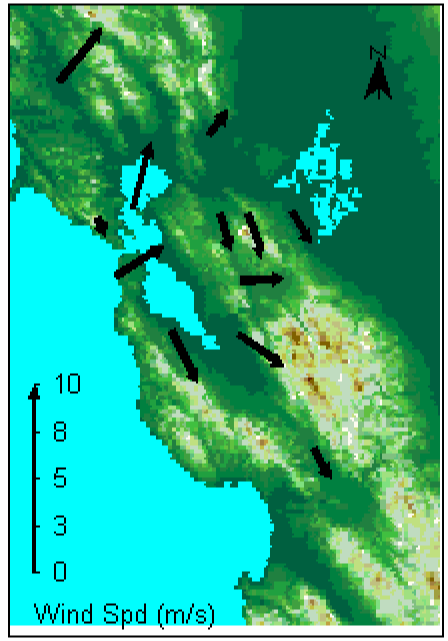
All 14 #4→#1 transitions are 3 days; $[O_3]$ peaks on middle day.

Very high $[O_3]$; explains 7 of 20 multiday SFBA exceedances.

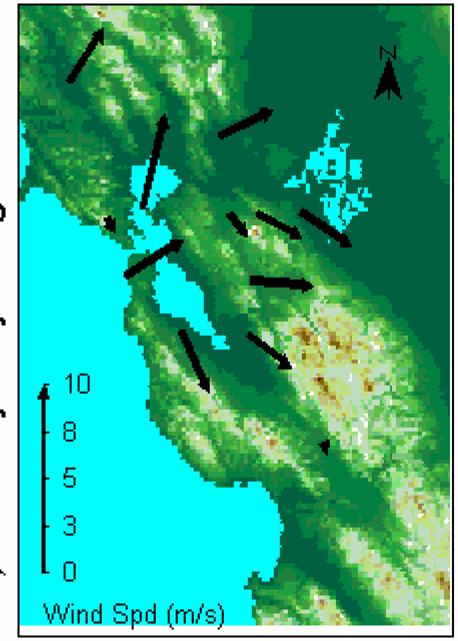
#4 → #1 Transition Wind Field



a) Last Day Fully Assigned to #4



b) #4 → #1 Transitional Day



c) First Day Fully Assigned to #1

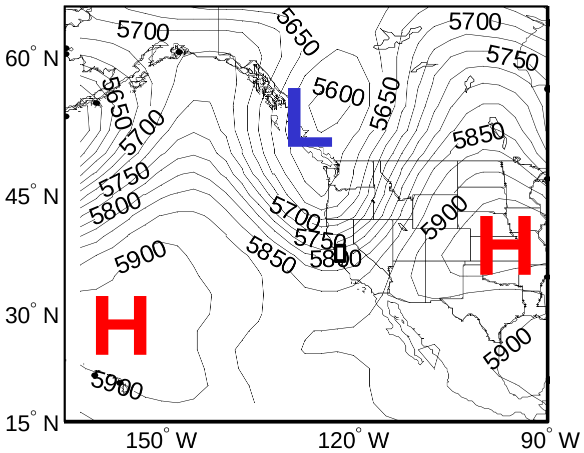
Shift in wind field from northerly to westerly over 3 days.

SCV seabreeze diminishes; LV seabreeze increases.

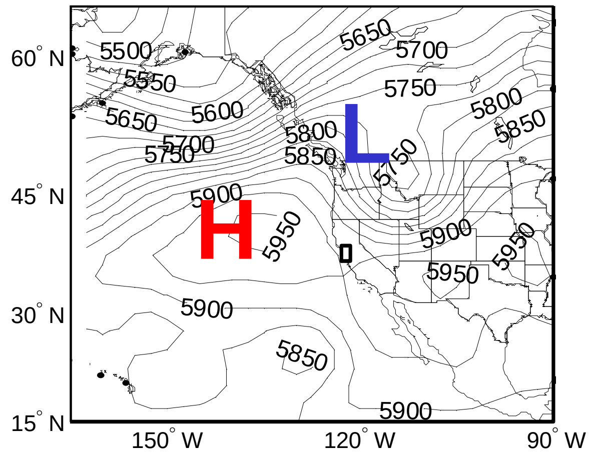
Location of peak daily ozone shifts from Santa Clara Valley to Livermore Valley.

#4 → #1 Transition Synoptic State

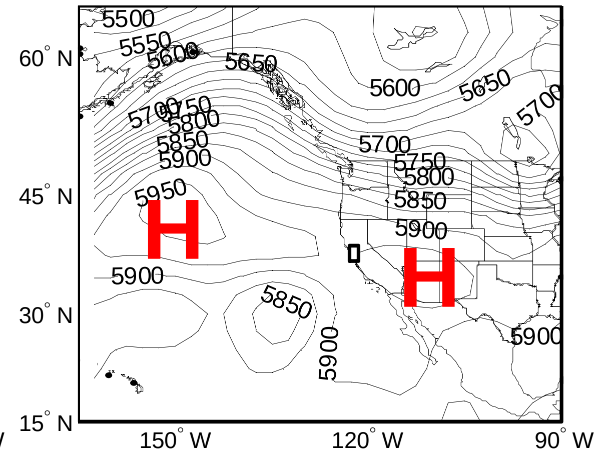
1800 UTC 05 Aug 2002



1800 UTC 08 Aug 2002



1800 UTC 10 Aug 2002



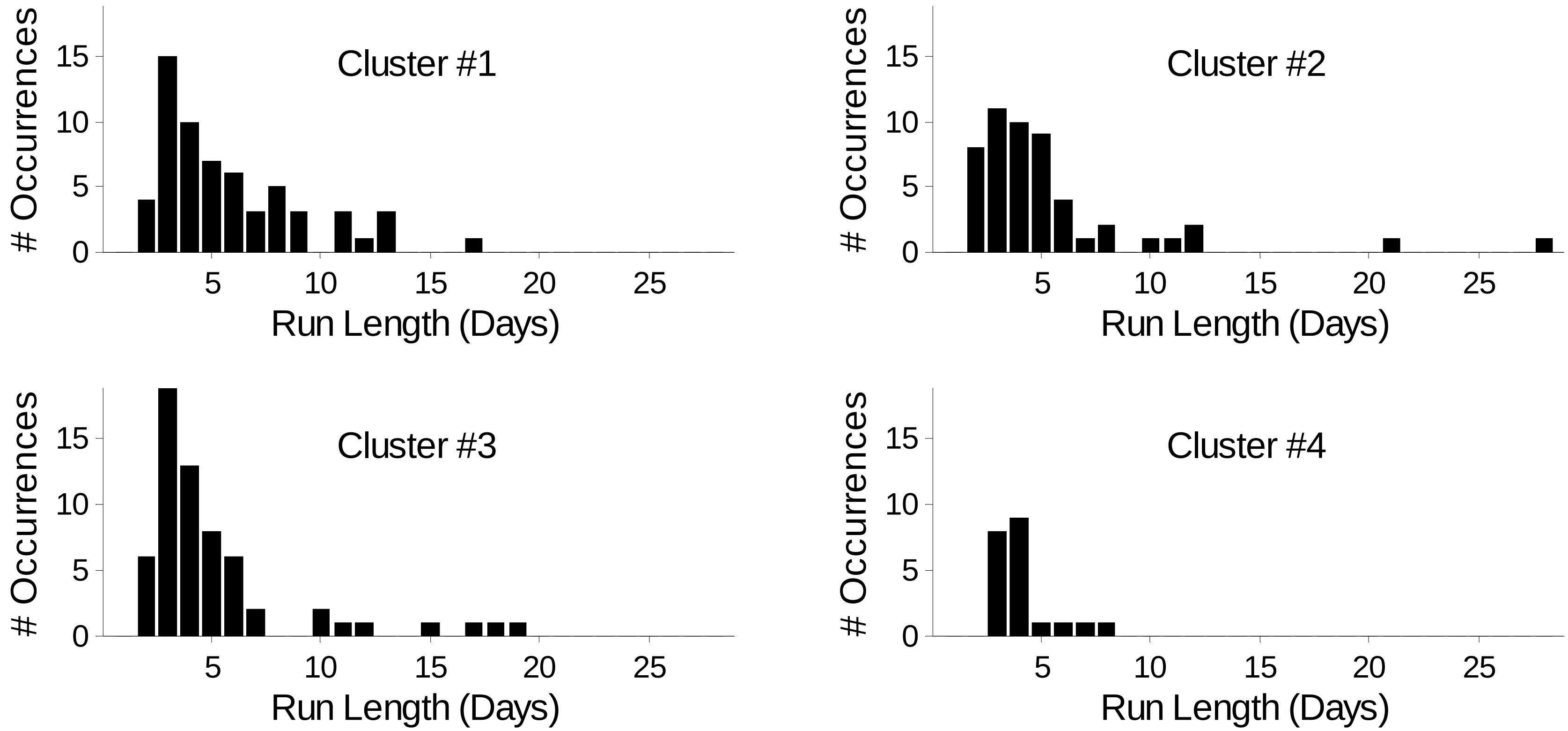
5 Aug: #2 (trough) realized. Offshore high pressure over Hawaii does not yet affect SFBA winds.

8 Aug: #4 realized as offshore high approaches SFBA and displaces trough inland. Northerly winds prevail.

9 Aug: #4→#1 transitional day (not shown).

10 Aug: #1 realized. Wide east-west band of high pressure conducive to multiday episodes.

Cluster Persistency



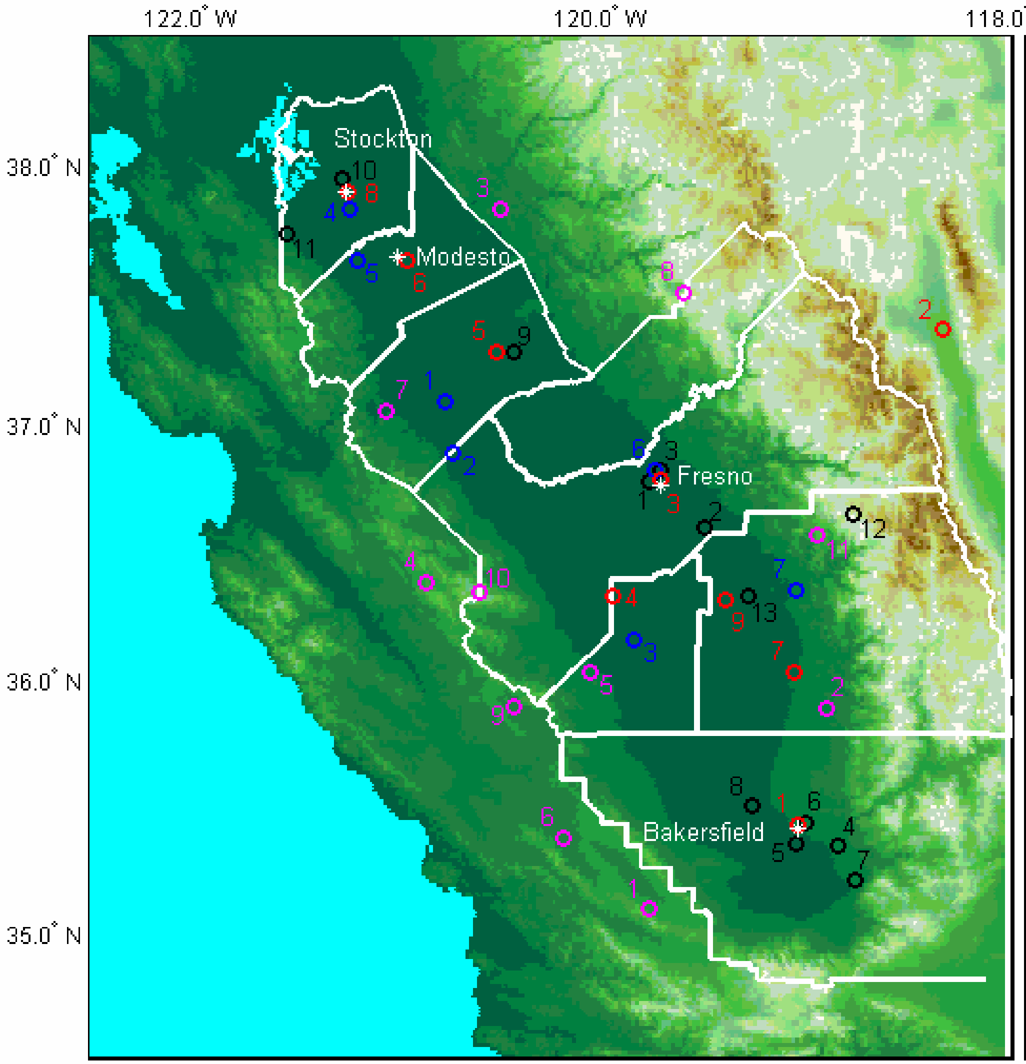
#1 is most persistent cluster. Onshore high pressure is stable.

#4 is least persistent cluster; realized for only 3—4 days. Offshore high pressure cell is unstable.

Synoptic Conceptual Model

- Offshore high (#4) usually transitions rapidly to onshore high (#1)
  - Northerly shift in SFBA winds indicates prolonged episode is imminent
  - Reverse onshore→offshore does not occur
- Onshore high pressure (#1) very stable
  - Requires low pressure of sufficient strength (#3) to displace #1
  - Bulk of episodes aside from #4→#1 transition
- Transitions from trough patterns (#2 and #3) are random
  - Little predictive capability from trough patterns
  - Evolution determined by prevailing conditions

San Joaquin Valley Met Networks



AIRS Stations:

1 FRESNO - FIRST STREET
2 PARLIER
3 CLOVIS-N VILLA AVENUE
4 EDISON
5 BAKERSFIELD-CALIFORNI
6 OILDALE
7 ARVIN-BEAR MOUNTAIN B
8 SHAFTER
9 MERCED - COFFEE ROAD
10 STOCKTON-HAZELTON
11 TRACY-PATTERSON RD.
12 SEQUOIA NATIONAL PARK
13 VISALIA

CIMIS Stations:

1 LOS BANO
2 PANOCHE
3 STRATFOR
4 MANTECA
5 MODESTO
6 FRESNO S
7 LINDCOVE

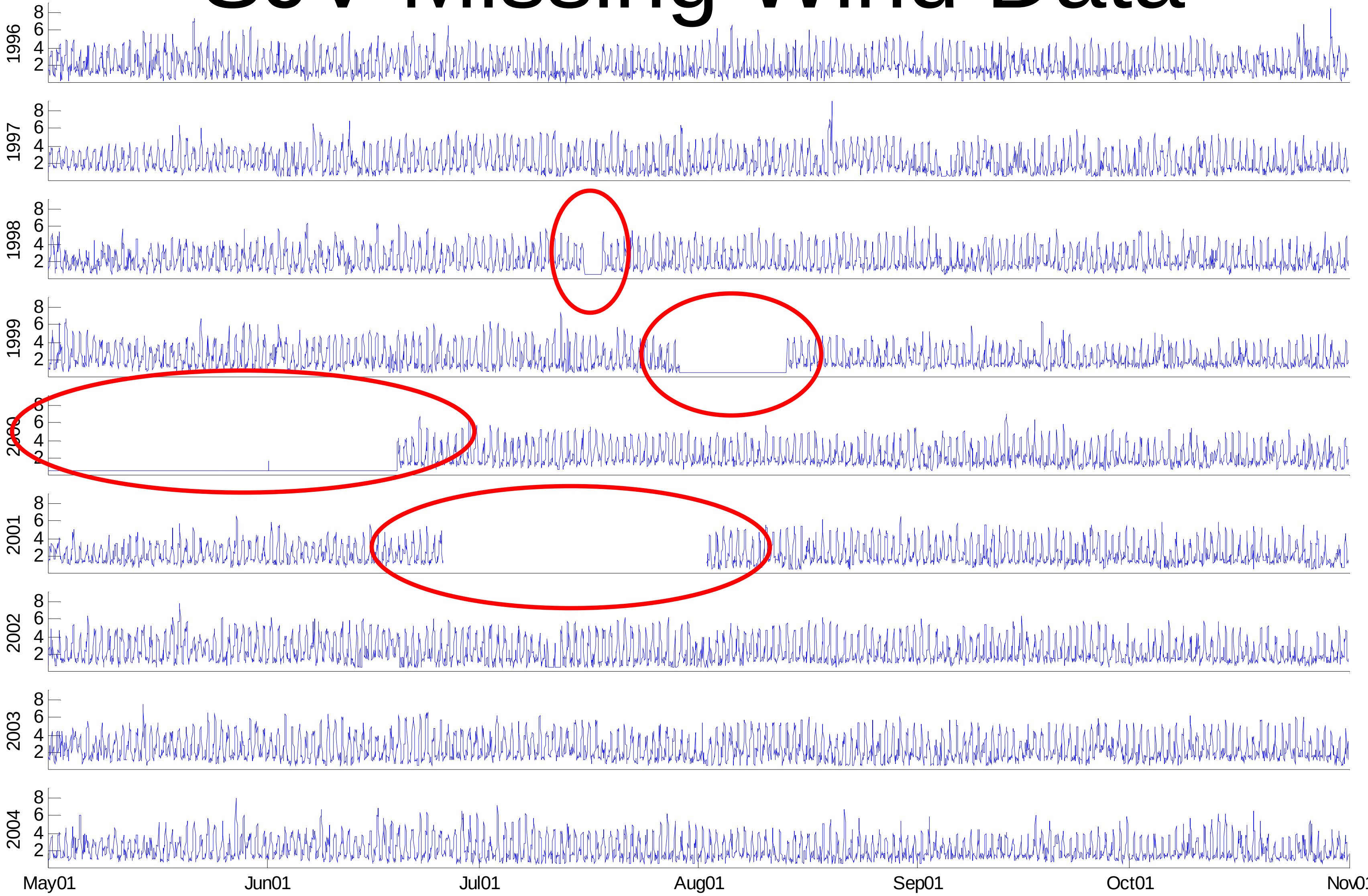
NCDC Stations:

1 BAKERSFIELD MEADOWS F
2 BISHOP AIRPORT
3 FRESNO YOSEMITE INTL
4 LEMOORE REEVES NAS
5 MERCED/MACREADY FLD
6 MODESTO CITY-COUNTY A
7 PORTERVILLE (AWOS)
8 STOCKTON METROPOLITAN
9 VISALIA MUNI (AWOS)

RAWS Stations:

1 CARRIZO
2 FOUNTAIN SPRINGS
3 GREEN SPRING
4 HERNANDEZ
5 KETTLEMAN HILLS
6 LA PANZA
7 LOS BANOS
8 MARIPOSA GROVE
9 PARKFIELD
10 SANTA RITA
11 SHADEQUARTER

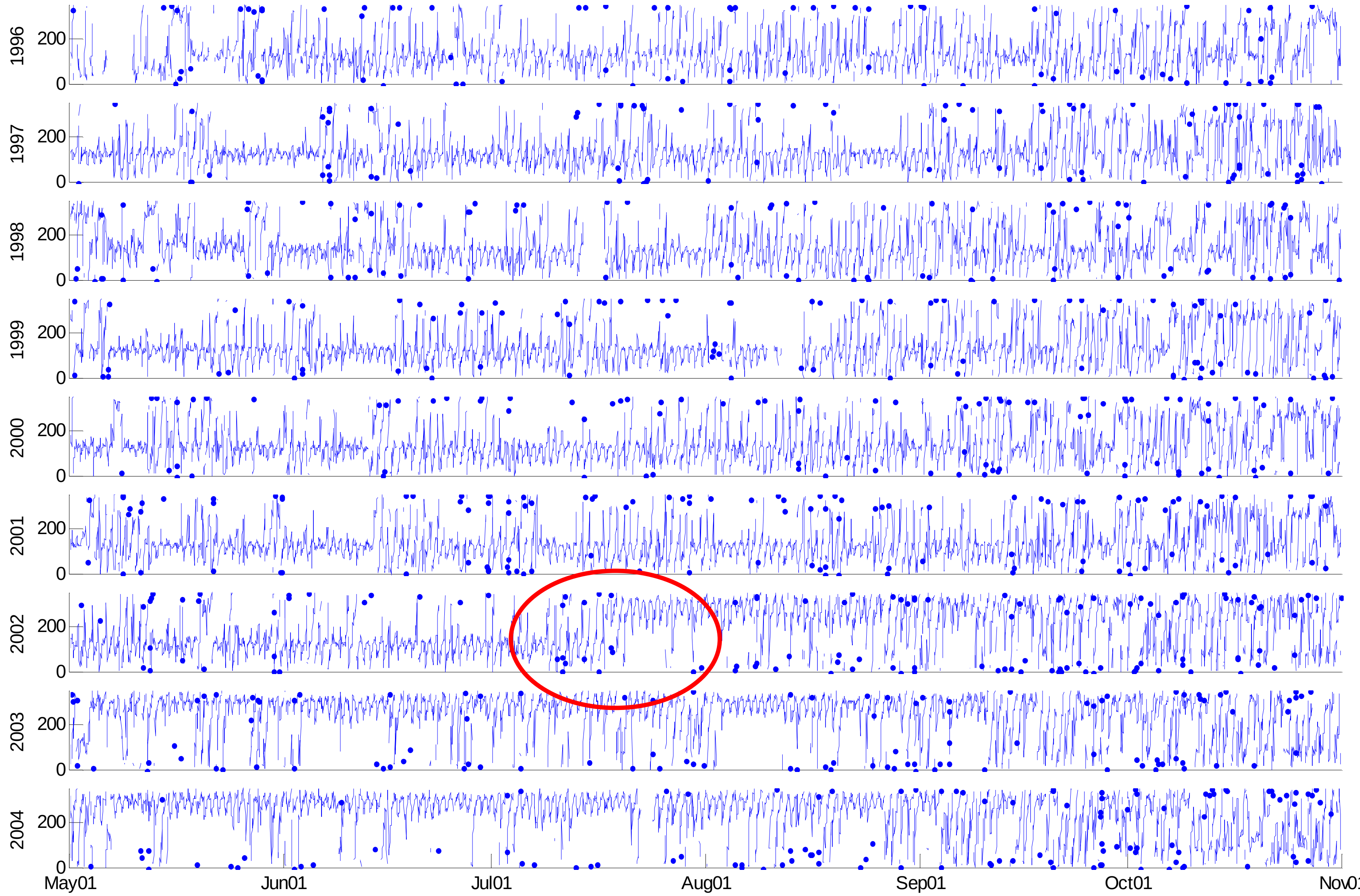
SJV Missing Wind Data



CIMIS station at Cuyama

Department of Chemical Engineering and Materials Science

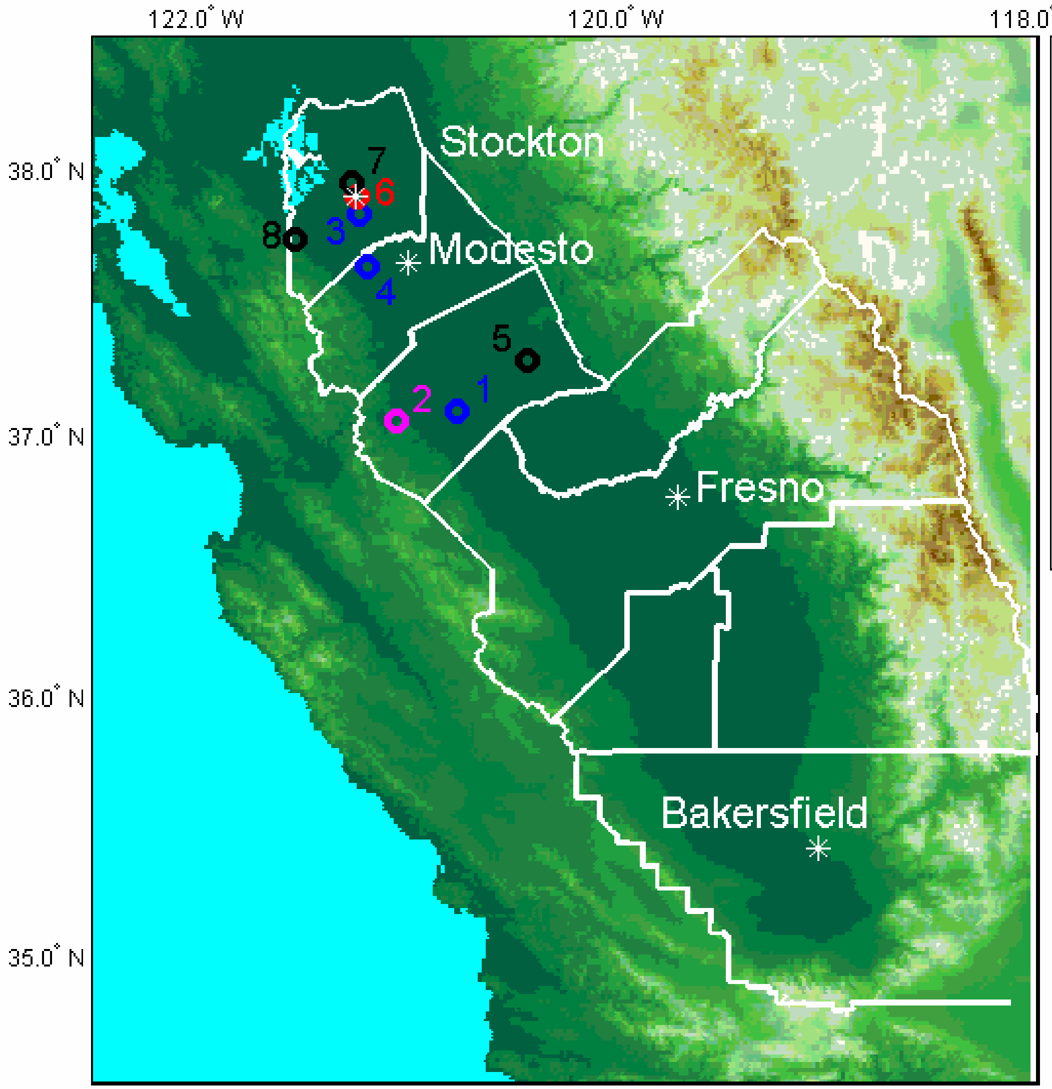
SJV Erroneous Wind Data



RAWS station at Green Spring

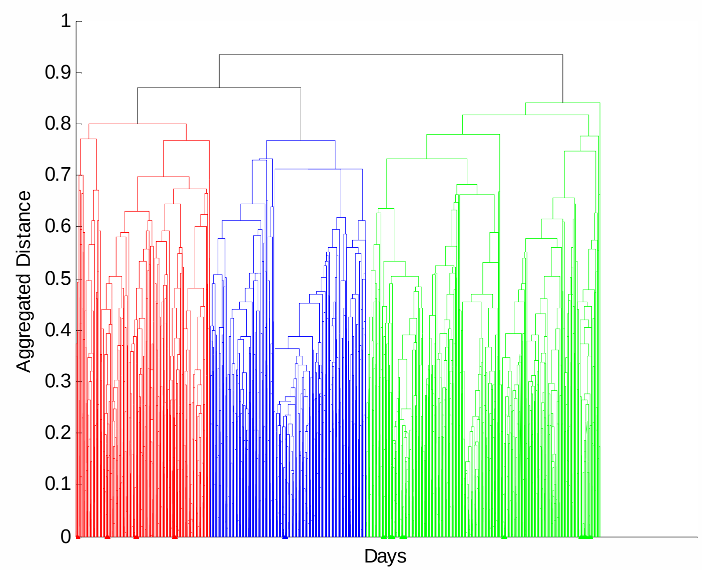
Department of Chemical Engineering and Materials Science

North SJV Wind Field Clustering



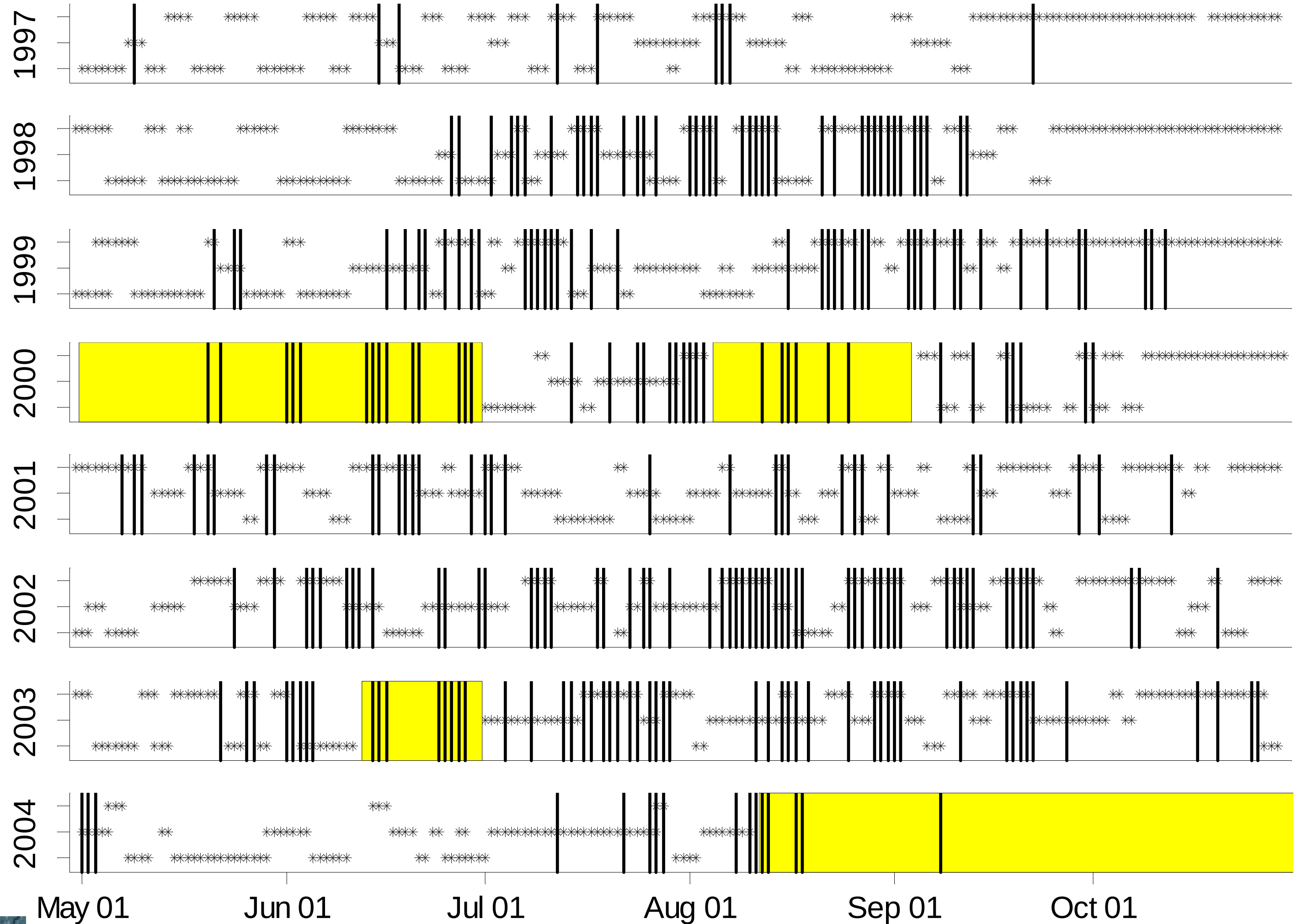
North SJV wind monitors:

1	LOS BANO	CIMIS
2	LOS BANOS	RAWS
3	MANTECA	CIMIS
4	MODESTO	CIMIS
5	MERCED - COFFEE ROAD	AIRS
6	STOCKTON METROPOLITAN	NCDC
7	STOCKTON-HAZELTON	AIRS
8	TRACY-PATTERSON RD.	AIRS

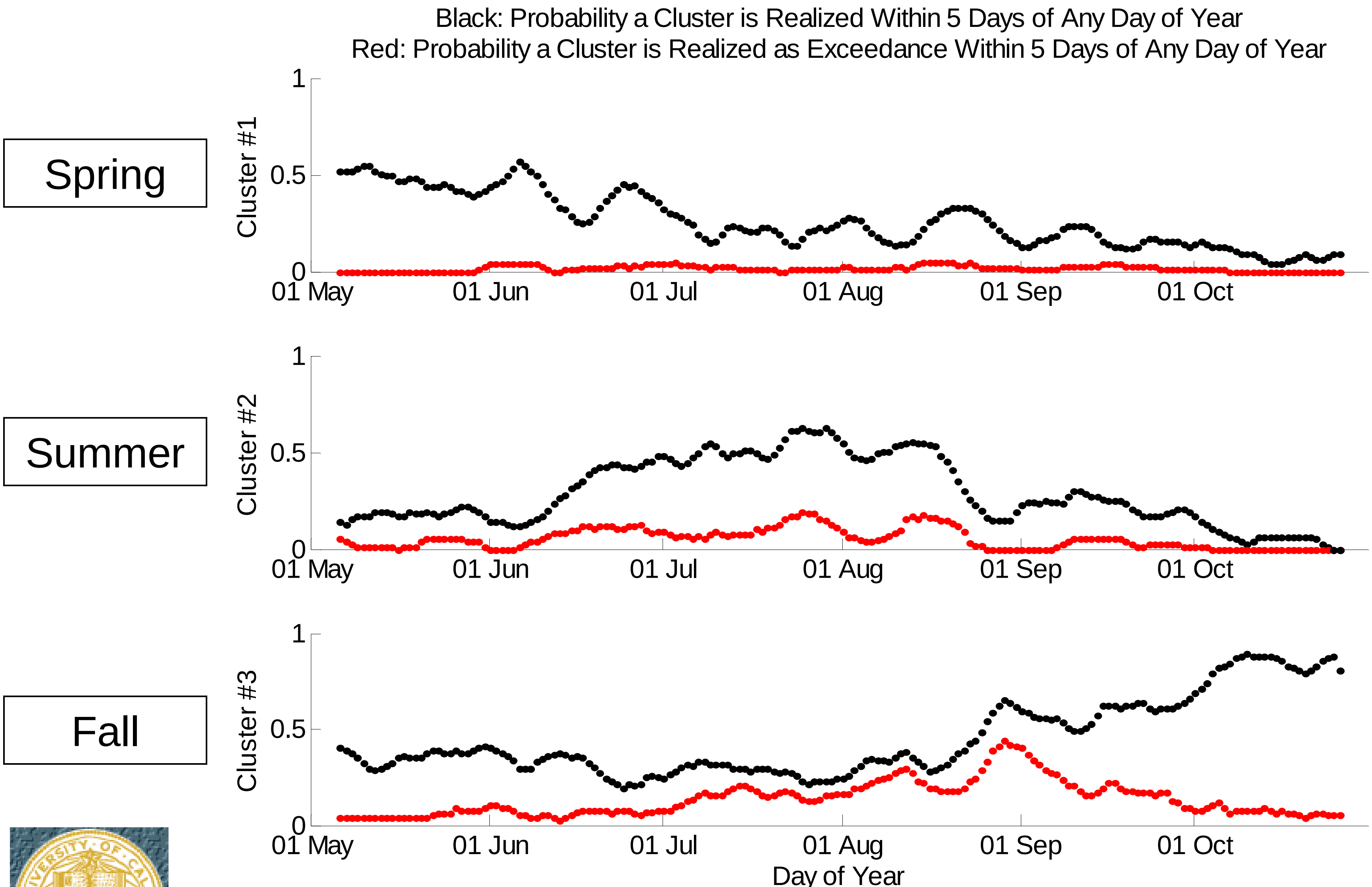


Time Series

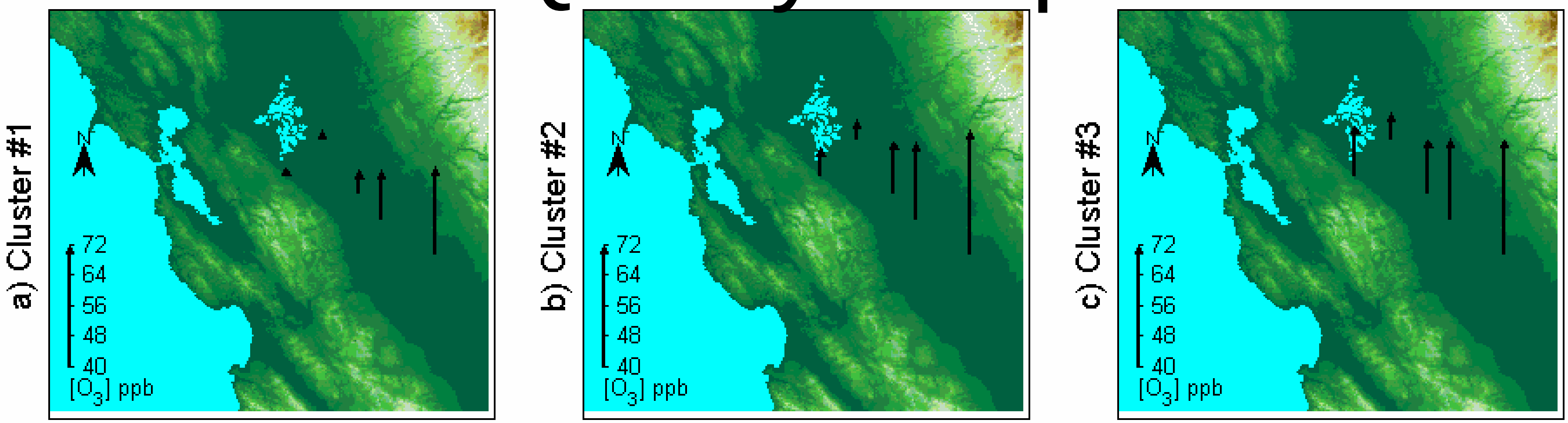
Daily Cluster Memberhip for 8 Summers of Wind Data and 8-hr Ozone Episodes.



Strong Seasonality



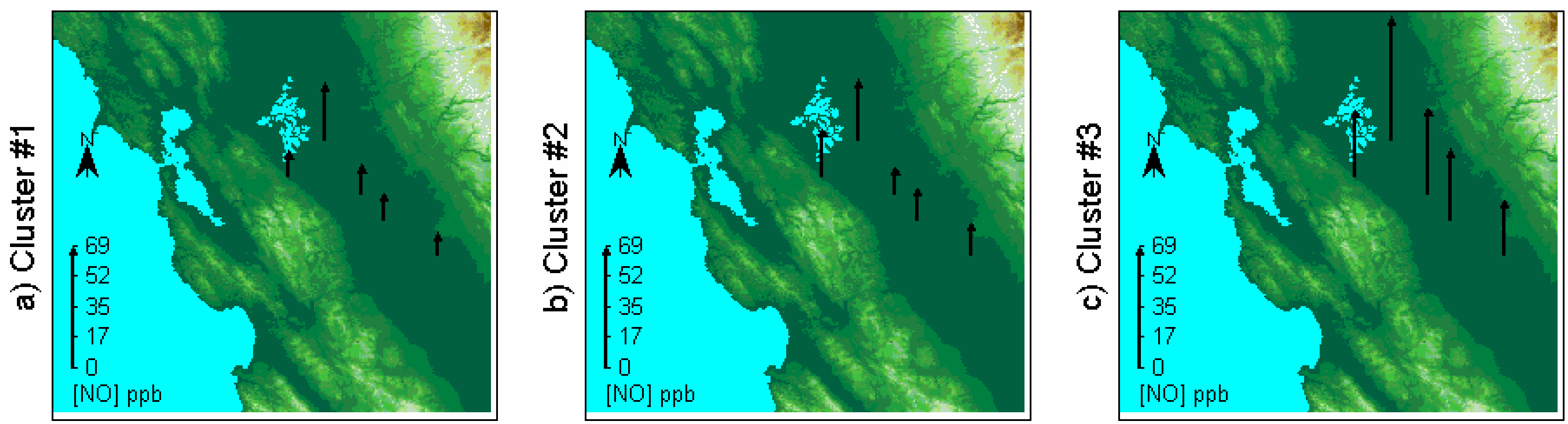
Air Quality Response



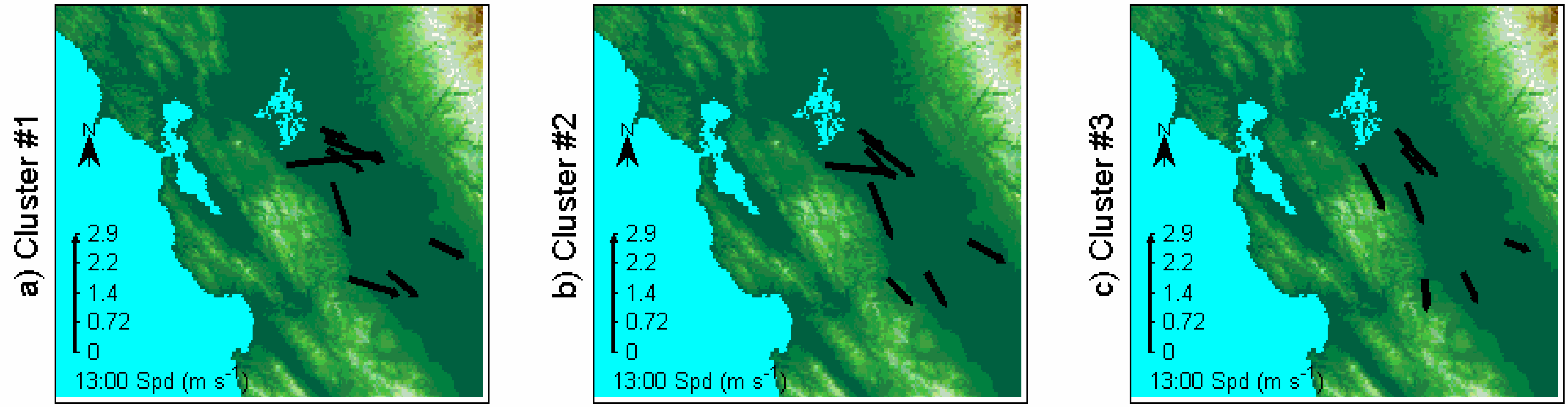
358 days, 22 episodes

398 days, 76 episodes

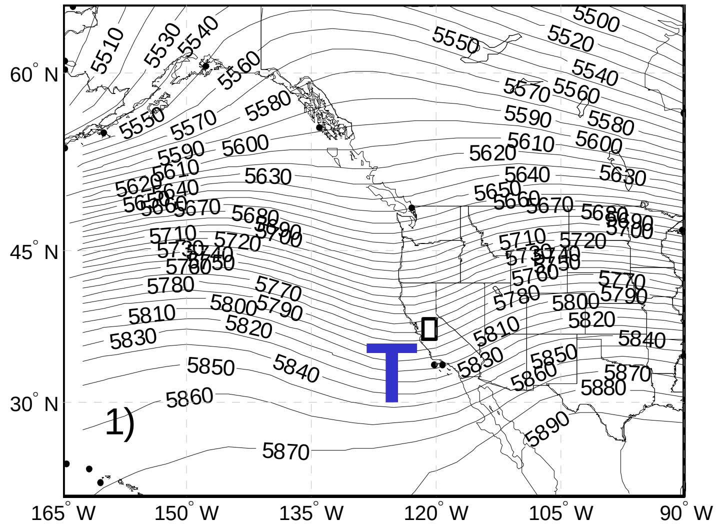
598 days, 170 episodes

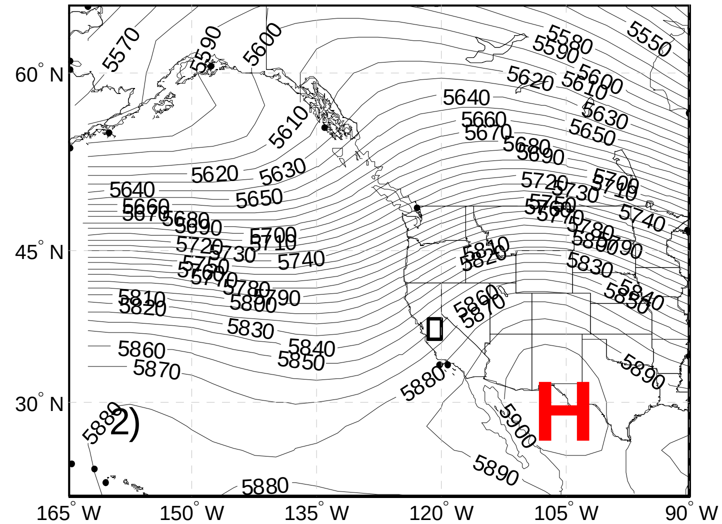


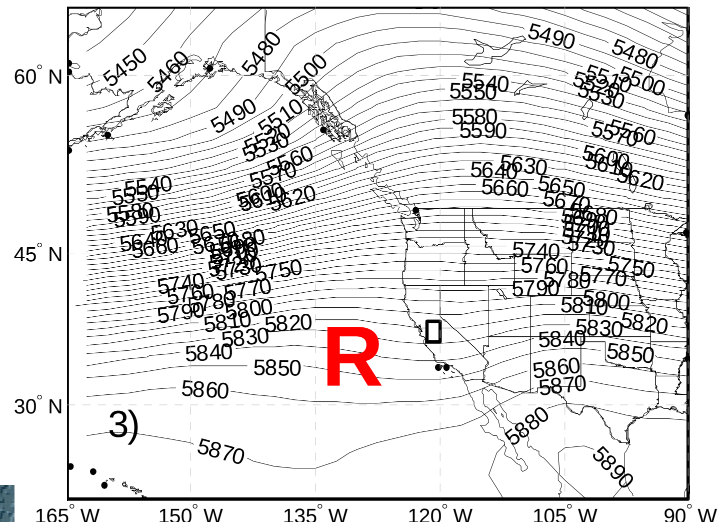
Midday Mesoscale Flow Patterns



Synoptic Influences







Same basic synoptic features as for SFBA:

#1 Trough

#2 Onshore high pressure

#3 Offshore ridge

Subclusters

- Each of 3 seasonal clusters break into several meaningful subclusters
  - Subclusters better explain regional O_3 variability
- Seasonality plays larger role for SJV than SFBA
  - Ex. Onshore high pressure for summer vs. fall produces different mesoscale flow pattern
  - Seasonal heating effects for SJV ???

Future Project Work

- Complete intra-basin analysis for 5 remaining basins
  - Confirm SJV met data set
  - Obtain met data for Sacramento Valley, Mountain Counties
- Perform inter-basin analysis
  - Effect of synoptic state on multiple basins
  - Time lags for air quality response
- Pre-1996 historical years?
  - Data availability issues

Recommendations

- Trend Analysis
  - Response of ozone to varying emissions levels under similar meteorological conditions
  - Annual meteorological variability such as El Nino
  - Wildfire effects on ozone
  - Incorporate VOC data for available days
- Refined analysis for small set of episodes
  - As suggested by ARB or CCOS staff
- Statistical analysis of sounding data
  - Model synoptic influences explicitly
  - 3D patterns at mesoscale
- Particulate Matter ???
  - Similar winter study for CCOS domain